AD_____

(2)

GRANT NO: DAMD17-90-Z-0033

TITLE: PHYSICAL CHARACTERIZATION OF CLOSTRIDIUM BOTULINUM NEUROTOXIN GENES

PRINCIPAL INVESTIGATOR: Nigel P. Minton, Ph.D.

CONTRACTING ORGANIZATION: Public Health Laboratory Service
Centre for Applied Microbiology
and Research
Division of Biotechnology
Porton Down, Salisbury
Wiltshire, SP4 OJG, United Kingdom

REPORT DATE: October 1, 1993

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. Army Medical Research and
Development Command, Fort Detrick
Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1, 1993	Final Report (7/2/90-9/1/93)

4. TITLE AND SUBTITLE
Physical Characterization of Clostridium Botulinum Neurotoxin Genes

5. FUNDING NUMBERS
Grant No.
DAMD17-90-Z-0033

6. AUTHOR(S)
Nigel P. Minton, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Public Health Laboratory Service
Centre for Applied Microbiology and Research
Division of Biotechnology
Porton Down, Salisbury
Wiltshire, SP4 OJG, United Kingdom

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research & Development Command
Fort Detrick
Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A total of 4 structural genes (*bot*) encoding botulinum neurotoxins (BoNT) have been cloned and their entire nucleotide sequences determined. These are the genes of the proteolytic *Clostridium botulinum* strains Danish (type B), NCTC 11219 (type E) Langeland (type F), and 89G (type G). The encoded toxins comprise 1291, 1252, 1278 and 1297 amino acids (aa), respectively. The determined BoNT/F sequence exhibits considerable divergence from that of a BoNT/F derived from a non-proteolytic strain of *C. botulinum* (ATCC 23387), and the BoNT/F produced by a strain of *Clostridium baratii* (ATCC 43756). Similar levels of divergence appear to exist between proteolytic and non-proteolytic type B neurotoxins. One consequence is that a vaccine based on the H_C polypeptide of a single representative of a particular serotype (notably types B and F) may not give protection against all members of that serotype. Attempts to impose control on a *C.acetobutylicum* expression system were unsuccessful. Gene fusions were constructed between a glutathione-S-transferase gene and *botA* DNA encoding various portions of the BoNT/A H-chain. No direct evidence for the production of recombinant protein in *C. acetobutylicum* was obtained.

14. SUBJECT TERMS
Foreign, BD, Clostridium, Cloning, Vaccines, BL3, Biotechnology, Gene expression, Gene cloning, DNA Transfer, Neurotoxin genes, Clostridium botulinum

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

NPM Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

NPM In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

NPM In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

NPM 1st October 1993
PI - Signature Date

CONTENTS

CONTENTS

Page Number

TABLES

FIGURES

SUMMARY

The structural genes (*bot*) encoding botulinum neurotoxin (BoNT) have been cloned from the *Clostridium botulinum* strains Danish (type B), NCTC 11219 (type E) Langeland (type F), and 89G (type G), and their nucleotide sequences determined. This has shown BoNT/B, BoNT/E, BoNT/F and BoNT/G to be respectively composed of 1291, 1252, 1278 and 1297 amino acids (aa), making the type E serotype the smallest characterised BoNT. Comparative alignment of translated aa sequences, and BoNT/A, C, D, and tetanus toxin (TeTx), demonstrates that clostridial neurotoxins are composed of highly conserved aa domains interspersed with aa tracts exhibiting little overall similarity. On the basis of aa similarity, TeTx is indistinguishable from a BoNT. In total 63 aa, out of an average 440, are absolutely conserved between L chains, and 93 out of 842 between H chains. The most divergent region corresponds to the carboxyterminus of each toxin, reflecting differences in specificity of binding to neurone acceptor sites. The relative order of relatedness varies according to which dichain component is compared. Recombinational events between different *bot* genes may therefore have taken place during evolution. The amino acid sequence of the BoNT/F determined in this study (isolated from a proteolytic *C. botulinum* strain) exhibits considerable divergence from that of a BoNT/F derived from a non-proteolytic strain of *C. botulinum* (ATCC 23387), and the BoNT/F produced by a strain of *Clostridium baratii* (ATCC 43756). Thus, the L- and H-chain of Langeland and ATCC 43756 share only 63% and 79%, respectively. Similar levels of divergence apparently exist between the neurotoxins of proteolytic and non-proteolytic type B *C. botulinum* strains. This order of divergence means that a vaccine based on the H_C polypeptide of a single representative of a particular serotype (notably types B and F) may not give protection against all members of that serotype.

Attempts to formulate genetic systems in *Clostridium sporogenes* were unsuccessful. Use was therefore made of an expression system, developed in this laboratory independently of this contract, for *Clostridium acetobutylicum*. Although the promoter in question (*fac*) is subject to regulatory control in *E.coli*, similar control could not be achieved in *C. acetobutylicum*. In the absence of a regulated system, attempts were made to effect the constitutive expression of *botA* subfragments in *C. acetobutylicum*. To aid in the subsequent purification of recombinant polypeptides, a strategy was formulated whereby they would be produced as a fusion protein with glutathione-S-transferase (GST), whose encoding gene exhibits a similar codon usage to clostridial genes. To accomplish this, DNA encoding the H_C fragment of BoNT/A (aa 855 to 1296) was fused to the extreme 3'-end of the GST gene, using PCR methodologies. To ensure eventual translation of the transcribed gene fusion in a Gram-positive host, a synthetic sequence specifying the ribosome binding site (RBS) of the TeTx gene was positioned immediately 5' to the translational start codon of the GST gene. The completed gene fusion was placed under *fac* transcriptional control by its insertion into pMTL500F. No evidence for the production of a recombinant protein was obtained when Western blots were performed on the lysates of *E. coli* cells carrying the resultant plasmid, pGAC501F, using either anti-BoNT/A or anti-GST antibody. Although cells carrying pGAC501F produced abnormal amorphous growth on solidified media, no evidence for the presence of inclusion bodies was forthcoming. Plasmid pGAC501F was subsequently found to be incapable of transforming either *B. subtilis* or *C. acetobutylicum*, a consequence, it is believed, of the production of the desired fusion protein. Derivative plasmids of pGAC501F were constructed in which the region encoding the entire BoNT/A H_C fragment was replaced with *botA* DNA encoding the NH_2- or COOH-terminal half of the H_C fragment (plasmids pGAC503F & pGAC504F, respectively). These new plasmids were now able to transform both Gram-positive hosts. The presence of a novel fusion protein could not, however, be detected in the lysates of transformed cells. Preliminary experiments, involving placement of the Fd RBS immediately 5' to the GST start codon, suggest that the TeTx RBS may be responsible for the lack of detectable protein.

INTRODUCTION

1. NATURE OF THE PROBLEM

The often fatal condition of botulism is caused by a group of highly toxic proteins (botulinum neurotoxin, BoNT) produced by certain species of clostridia, principally *Clostridium botulinum* (Sugiyama, 1980). On the basis of their serological properties, seven distinct types of BoNT are recognised, and have been designated BoNT/A to G. They exert their effects on vertebrates by blocking the release of the neurotransmitter acetylcholine in presynaptic nerve termini, resulting in neuromuscular paralysis (Habermann and Dreyer, 1986; Simpson, 1989). Although BoNT is synthesised as a single polypeptide chain (M_r approximately 150, 000), proteolytic cleavage generates the more toxic dichain form, in which a 50 000 Da polypeptide light (L) chain and a 100 000 Da heavy (H) chain are linked by a disulphidryl bridge. The different types of *Clostridium botulinum* exhibit differential efficiencies in nicking of the single chain to the dichain form. Thus, BoNT/A exists principally as a dichain, BoNT/B exists as a mixture of predominantly single chain with some dichain, whereas BoNT/E is found essentially only in the single chain form (Dasgupta, 1990). Purified single chain toxin may be converted to the dichain form *in vitro* by proteolytic cleavage with trypsin (Dolly et al., 1984).

The overall structure and mode of action of BoNT is shared by a second clostridial toxin, namely tetanus (TeTx) of *Clostridium tetani* (Welloner, 1982). They differ in that whereas BoNT acts at the nerve periphery, TeTx blocks the release of inhibitory amino acids in the central nervous system. The neuroparalytic action of both types of neurotoxin has been suggested (Simpson, 1986) to be composed of three distinct phases: (i) binding of the toxin to neurone acceptor sites; (ii) an energy-dependent internalisation stage in which the toxin, or part of it, enters the nerve cell, and; (iii) the eventual blockade of neurotransmitter release. Although the exact mechanisms involved remain poorly understood, it is generally assumed that the L chain possesses the pharmacological activity (Bittner et al., 1989; Ahnert-Higler et al., 1989) and the H chain is responsible for binding of the dichain to cell surface acceptors and thereafter internalisation through the cell membrane (Simpson , 1989). Some evidence has been obtained suggesting that the channel forming activity resides in the NH_2-terminal portion of the H chain (Mochida et al., 1989; Poulain et al., 1990) and acceptor recognition sites in the COOH-terminus (Morris et al., 1981; Shone et al., 1985; Kozaki et al., 1987; 1989).

The effectiveness of modern food-preserving processes in Western countries has made

outbreaks of botulism extremely rare. The frequent use of *C.botulinum* as a test organism in the food industry, and the growing use of the toxin by neurobiochemists, has, however, led to the development of human vaccines. The formulation of these vaccines has changed little since the early 1950s; partially purified preparations of the neurotoxins are toxoided by formaldehyde treatment and absorbed onto precipitated aluminium salts. Using such methodology, polyvalent vaccines (against ABCDE or ABEF) for human immunisation are currently available. Such vaccines suffer from the drawback of low immune response and considerable batch to batch variation due to the high proportion (60-90%) of contaminating proteins in toxoid preparations. Recent work has therefore concentrated on the development of procedures for the purification of toxins to near-homogeneity. This has been achieved with all but type G toxin (Shone *et al.* 1985; Evans *et al.*, 1987; Schmitt *et al.*, 1986). The use of purified toxins in the production of vaccines, however, suffers from the drawbacks of having to produce them under high containment and requires the presence of low levels of formaldehyde to prevent possible reversion of the toxoid to the active state.

2. BACKGROUND OF PREVIOUS WORK

Production of subunit vaccines have been investigated by a number of laboratories. In general, individual toxin subunits produce poor immune responses. A non-toxic fragment comprising the L-chain and the N-terminal portion of the H-chain (analogous to the AB fragment of tetanus toxin) of type A toxin has been shown to produce an immune response in guinea pigs comparable to the entire toxin (Shone and Hambleton, 1989). It has therefore been argued that production of such a toxoid polypeptide by recombinant means provides an excellent candidate for future vaccines. This would most simply be achieved by insertion of the appropriate coding sequences into specialised bacterial vectors, which then direct the expression of high levels of the protein in suitable bacterial hosts. The unparalleled sophistication of recombinant procedures and vectors of *E.coli* has resulted in this enterobacteria being the organism of choice in such processes. There are, however, a number of factors which suggest that *E.coli* is not the best candidate for undertaking the expression of clostridial toxin genes.

Although clostridial genes are reported to express moderately well in *E. coli* (reviewed by Young et al., 1989), this finding only applies to genes isolated from mesophiles encoding proteins substantially smaller (c. 30-40, 000 Da) than BoNT, or thermophilic genes (eg., from *C.thermocellum*) whose G + C content closely matches that of *E. coli*. Attempts to express clostridial genes encoding large polypeptides have met with either very limited success (eg., type A toxin of *Clostridium difficile*; von Eichel-Streiber, 1989) or total failure (eg., the bacteriocin of the *Clostridium perfringens* plasmid pIP401, Garnier and Cole, 1988). More

germane to BoNT have been the attempts to obtain expression of polypeptide fragments of TeTx. In the study of Eisel *et al.* (1986) various subfragments of the gene were expressed in *E. coli*, either initiating from the tetanus ATG or as fusion proteins. The levels attained were extremely poor, and it was concluded that no clone "led to the synthesis of sufficient amounts of toxin-specific protein to allow biological studies. At present these considerations argue against a large-scale production of toxoid based on genetically engineered non-toxic derivativc." Similar results were obtained by Fairweather *et al.* (1986, 1987), who expressed the C-terminal portion of the toxin (43% of the molecule) to levels less than 1 % of the cell's soluble protein. More recently, attempts to express subfragments encoding either the L-chain or substantial portions of the H-chain of the type A gene have met with little success (A.H. Bingham, personal communication). A further difficulty encountered in all these studies was considerable degradation of the polypeptides produced, even in protease minus *E.coli* hosts.

The reasons for the observed inefficient expression of large clostridial toxin genes would appear complex, but the apparent translational barrier is suggested (Eisel et al., 1986; Garnier and Cole, 1988) to be a consequence of the extremely biased codon usage exhibited by clostridial genes. Thus genes isolated from *Clostridium spp.* whose genomic DNA is of a high A+T content (greater than 70% A+T), exhibit an extremely strong discrimination against all degenerate codons ending in C or G, or, in the case of Ser and Arg, beginning with C. In the case of the neurotoxin type A gene (Thompson et al., 1990), 86.1% of Arg codons conform to AGN rather than CGN, 69% of Leu codons conform to UUA as opposed to CUN, while overall, 90.3% of the degenerate codons end in A or U. In the tetanus toxin gene the equivalent respective figures are 92.1%, 69.3% and 92.9%. A consequence of this codon bias is that many of those codons known to act as modulators of gene expression in *E.coli* (Grosjean and Fiers, 1982) occur extremely frequently in clostridial genes. eg., the type A neurotoxin gene exhibits a 53.8% preference for AUA (Ile), 43.7% preference for GGA (Gly) and an overall 86.1% preference of AGN (Arg) modulator codons. It would appear that although *E. coli* can tolerate a certain number of such codons, as occurs in genes of moderate size, the cumulative effect of the sheer volume of modulator codons present in clostridial neurotoxin genes results in a dramatic reduction in translational efficiency. The most logical solution to these problems would be to use a clostridial host, rather than *E.coli*.

3. PURPOSE OF THE PRESENT WORK

The production of a polyvalent vaccine against all known types of botulinum neurotoxins requires the availability of large quantities of pure protein which is; (i) capable of eliciting the production of neutralising antibody in humans and; (ii) non-toxic to personnel involved in its isolation, purification and formulation into a vaccine. These criteria cannot be currently met by producing authentic neurotoxin from natural clostridial strains. Although the desired

subunit vaccine could conceivably be produced by recombinant means, as discussed above, translational barriers suggest that *E. coli* cannot be employed as the recombinant host. A major objective of this study was therefore to develop a clostridial expression system, ideally based on a non-toxinogenic host closely related to *C.botulinum*, and test its utility by expressing various non-toxic polypeptides (principally derived from the H-chain moiety) of the type A neurotoxin. The immunogenicity of these recombinant polypeptides would then be evaluated as potential subunit vaccines. In parallel, the second principal objective has been to clone other neurotoxin genes (types B, E, F and G) and derive their complete primary amino acid sequences by nucleotide sequence analysis. Selected polypeptides of these neurotoxins could then also be produced using the recombinant host/vector system and their potential as subunit vaccines ascertained. At the end of these studies it was anticipated that a system for producing high levels of non-toxinogenic neurotoxin polypeptides will have been developed which may be used in the formulation of a general botulism vaccine against types A, B, E, F and G. Furthermore, the availability of the complete primary amino acid sequences of these toxins will facilitate future work which may be aimed at deriving vaccines based on synthetic peptides.

4. METHODS OF APPROACH

4.1 Development of a Clostridial Expression System

Our initial strategy was to choose a *Clostridium* sp., taxonomically closely related to *C. botulinum*, and formulate procedures for introducing recombinant DNA. Our choice as the host was *Clostridium sporogenes* (taxonomically considered to be a non-toxigenic species of *C. botulinum*; Cato and Stackebrandt, 1989) and the DNA transfer procedures investigated, electroporation and conjugative transfer. The former procedure, ubiquitous in its application to Gram-positive bacteria (see Chassy *et al.*, 1988; Lucansky *et al.*, 1988), relies on the transient introduction of pores into the cell membrane, by applying an electrical discharge across cell suspensions, through which exogenous DNA may pass. We have previously used electroporation for the successful introduction of plasmids into *C. acetobutylicum* (Oultram *et al.*, 1988a), and similar protocols have been published for *C. perfringens* (eg., Allen and Blaschek, 1988). Conjugative transfer relies on mobilisation of the cloning vector into *C. sporogenes* by intergeneric matings (Trieu-Cuot *et al.*, 1987). We have constructed a plasmid, pMTL30 (Williams et al., 1990a; 1990b), which carries the ColE1 replicon, the Gram-positive erythromycin (Em) resistance (^{R}) gene of pAMB1 (Brehm *et al.*, 1987), the *E.coli lacZ'*/multiple cloning region of pMTL20 (Chambers *et al.*, 1988), and the *oriT* region of plasmid RK2. Plasmid derivatives, in which the replication origins of either pCB101 (a *Clostridium butyricum* plasmid; Minton and Morris, 1981) or the streptococcal plasmid

pAMB1 have been inserted, have been shown to be mobilised from an *E.coli* donor to an *C. acetobutylicum* recipient at frequencies of up to 10^{-5} per donor (Williams et al., 1989a; 1989b). In out attempts to transfer DNA into various strains of *C. sporogenes* the plasmid vehicles utilised were endowed with the replicative origins of either pCB101 or pAMB1. The latter type of vector was preferred as it has proven to possess an extremely broad host range amongst Gram-positive bacteria, and exhibits a high degree of structural stability (Bruand et al., 1990; Swinfield et al., 1990). As it cannot be assumed that these replicons will function in *C. sporogenes* it was envisaged that replicons could be cloned from indigenous *C. sporogenes* cryptic plasmids into 3 different types of "in-house" Gram-positive replicon cloning vector (ie., plasmids only capable of replicating *in E.coli*). These vectors (pMTL20E, pMTL20C and pMTL20T) carry three different Gram-positive resistance genes (*erm*, *cat* and *tetP*, respectively), all of which have been shown to express in *Clostridium spp.* (see Minton and Oultram, 1988; Abraham and Rood, 1985).

Having formulated procedures for DNA transfer we proposed to endow constructed shuttle vectors with efficient transcription/ translation signals to facilitate high expression of appropriately inserted heterologous genes. Since ribosomal RNA (rRNA) operons are generally transcribed efficiently it was proposed that the rRNA genes of *C. sporogenes* would be the source of transcriptional initiation and termination signals. Once cloned and characterised the identified promoter region was to be modified by advanced genetic engineering (ie., creation of restriction sites by site-directed mutagenesis and insertion of required sequences as synthesised "units") to create an expression cartridge. This would consist of a portable restriction fragment, carrying (in sequential order): the rRNA promoter -35 and -10 elements; a synthetic *E.coli lacZ* operator sequence positioning immediately following the rRNA +1; a synthetic ribosome binding site (SD) complementary to the determined *C. sporogenes* 16s RNA; at an appropriate distance from the SD, a recognition sequence for *Nde*I (CATATG), followed by the *lacZ'*/multiple cloning sites of plasmid pMTL20, whereby the ATG represents the translational initiation codon of *lacZ'*; finally the *lacZ'* region would be followed by the transcriptional termination signals of the rRNA operon. The efficiency of the system could be tested using a suitable promoter-less reporter gene, eg., *cat*. The presence of the *lacZ* operator site should allow repression of expression during construction in *E.coli* (by the presence of the high copy number *lacI*^{q} plasmid pNM52. Gilbert *et al.*, 1986), and thereafter regulated expression of the gene in clostridia. It was envisaged that this could be achieved in an analogous fashion to that used in *B.subtilis* (Le Grice *et al.*, 1987), where a plasmid borne copy of the *lacI*^{q} gene is placed under the transcriptional control of a moderate clostridial promoter (we will use the *Clostridium pasteurianum leuB* promoter, cloned and sequenced in this laboratory), and induction of the rRNA expression cartridge elicited by addition of IPTG.

Our subsequent failure to elicit demonstrable DNA transfer to any of the strains of *C.*

sporogenes tested necessitated a substitution of the intended recombinant host with *C. acetobutylicum*. This clostridia has a number of advantages over *C. sporogenes*. On a practical level, we have already developed the necessary means of manipulating this species. Equally as important, this species has no known association with human disease and should therefore command a lower Access factor in any proposed recombinant experiments. The proposed expression of BoNT gene subfragments can therefore be undertaken at a lower category of containment. Furthermore, parallel studies undertaken in this laboratory have resulted in the construction of an expression cartridge, similar to that described above, based on the promoter of the ferredoxin (Fd) gene of *Clostridium pasteurianum*. This promoter, modified by the insertion of the *E. coli lac* operator, has been designated the *fac* promoter and shown to direct the expression of a *cat* gene in *C. acetobutylicum* NCIB 8052 to between 5 and 10% of the cells' soluble protein. Once *C. sporogenes* was abandoned as the recombinant host, efforts were therefore switched to attempting to obtain *lacI* expression in NCIB 8052.

4.2 Cloning of Botulinum Neurotoxin Genes:

The strategies utilised in the cloning of the type B, E, F and G neurotoxin genes were devised to minimise the risk of obtaining a toxinogenic *E. coli* recombinant clone, and mirrored the measures taken in the cloning of the BoNT/A gene, *botA* (Thompson et al., 1990). Thus, as both L and H chain are required for toxicity (Simpson, 1989), only DNA fragments encoding principally one component of the dichain were cloned. Where genomic fragments were cloned, their coding potential was determined by the construction of genomic maps using *botA* DNA probes in Southern blots. Furthermore, they were always isolated by two-stage agarose gel size fractionation to minimise the risk of cloning contiguous DNA fragments. As more nucleotide sequence information became available, specific regions were amplified for cloning by polymerase chain reaction (PCR).

BODY

1. CLONING OF THE BoNT GENES

1.1 MATERIALS AND METHODS

Bacterial strains, plasmids and culture conditions.

The source of chromosomal DNA was *C. botulinum* strain B/Danish, the type E strain NCTC 11219, the type F Langeland strain and the type G strain 89G. The recombinant host used for cloning experiments *E. coli* TG1 Δ[*lac-pro*] *supE thi hsdD5*/ F'- *traD36 proA*$^+$ *B*$^+$ *lacI*^{Q} *lacZΔM15*). Cloning vectors employed were plasmids pMTL32 (this study), pMTL20 (Chambers et al., 1988), pCR1000 (Mead et al., 1991), and the M13 phages mp18 and mp19 (Yanisch-Perron, 1985). *C. botulinum* was cultivated in USA II broth (2% peptone, 1% yeast extract, 1% N-Z amine, 0.05% sodium mercaptoacetate, 1% glucose, pH 7.4), and *E. coli* in L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl). Solidified medium (L-agar) consisted of L-broth with the addition of 2% (w/v) agar (Bacto.Difco). Antibiotic concentrations used for the maintenance and the selection of transformants were 50 μg/ml ampicillin (pMTL32/ pMTL20) and 50 μg/ml kanamycin (pCR1000). Restriction endonucleases and DNA modifying enzymes were purchased from Northumbria Biochemicals Ltd, Taq polymerase from United States Biochemical Corporation and radiolabel from Amersham International.

Purification and manipulation of DNA.

Transformation of *E. coli* and large-scale plasmid isolation procedures were as previously described (Minton et al., 1983). Small-scale plasmid isolation was by the method of Holmes and Quigley (1981), while chromosomal DNA from *C. botulinum* was prepared essentially as described by Marmur (1961). Restriction endonucleases and DNA modifying enzymes were used under the conditions recommended by the supplier. Digests were electrophoresed in 1% agarose slab gels on a standard horizontal system (BRL Model H4), employing Tris-borate-EDTA (0.09 M Tris borate, 0.002 M EDTA) buffer. Fragments were isolated from gels using electroelution (McDonnell et al., 1977). All primary cloning procedures were undertaken under United Kingdom ACGM C2 containment conditions, and total cell lysates of all

recombinants carrying cloned material were tested in mice for the absence of toxic polypeptides.

DNA/DNA hybridisation experiments.

DNA restriction fragments were transferred from agarose gels to "zeta probe" nylon membrane using the procedure of Reed and Mann (1985). After partial depurination with 0.25 M HCL (15 min), DNA was transferred in 0.4 M NaOH by capillary elution for between 4 and 16 hours. Bacterial colonies were screened for desired recombinant plasmids by *in situ* colony hybridisation (Grunstein and Hogness, 1975), using nitrocellulose filter disks (Schleicher and Schull, 0.22 μm). The gel purified *botA* DNA fragments were labelled with [α-^{32}P] dATP using a multiprime kit supplied by Amersham International. Hybridisations were carried out as previously described (Thompson et al., 1990), at temperatures ranging from 45 to 60 °C.

Nucleotide sequence of bot plasmid inserts.

The nucleotide sequences of plasmid inserts were determined by a number of different strategies. In some instances the entire insert was excised, circularised by treatment with T4 ligase and size fractionated 500-1000 bp fragments generated by sonication cloned into the *Sma*I site of M13mp18 (for experimental conditions, see Minton et al., 1986). Approximately 50 templates were then sequenced by the dideoxynucleotide method of Sanger et al (1980) using a modified version of bacteriophage T7 DNA polymerase, "sequenase^{R}" (Tabor and Richardson, 1987). Experimental conditions used were as stated by the supplier (United States Biochemical Corp.). The inserts of other plasmids (eg., pCBB2 and pCBB3) were sequenced using templates derived by subcloning the entire region between the appropriate sites of M13mp18 and M13mp19. Sequence data obtained employing universal primer was then sequentially extended by the use of custom-synthesised oligonucleotide primers. In certain instances, templates were generated by the insertion of *Dra*I restriction subfragments into the *Sma*I site of M13mp18. In all cases the sequence was determined on both DNA strands. On some occasions PCR amplified DNA was cloned directly into either pCR1000 or ddT-tailed, *Sma*I cut M13mp18 (prepared by incubating *Sma*I cut DNA with terminal transferase in the presence of dideoxy TTP), and the resultant plasmid/ template sequenced with universal primer. DNA sequence data was analysed using the computer software of DNASTAR Inc.

Amplification of DNA by PCR.

Amplification of *C. botulinum* DNA was undertaken by polymerase chain reaction (PCR), using an M J Research Inc. Thermal cycler. Reaction mixtures comprised, 10 mM Tris-HCl, 50 mM KCl, 3 mM $MgCl_2$, 0.1 mM dNTP, 30 nmol of each primer, 2.5 units of Taq polymerase, and 10 ng of *C. botulinum* genomic DNA, in a final volume of 0.1 ml. Amplification was for 30 cycles, as follows: 1.5 min at 93°C; 3 min at 37°, and; 3 min at 72°C. For inverse PCR, 140 ng of chromosomal DNA, cleaved with an appropriate restriction endonuclease, was ligated overnight at 14°C in a 50 μl volume and a 10 μl portion of the resultant concatenated DNA used in PCR.

1.2 CLONING/ SEQUENCING OF THE BoNT/E GENE

1.2.1 Summary

The entire structural gene of the *Clostridium botulinum* NCTC 11219 type E neurotoxin gene has been cloned as 5 overlapping DNA fragments, generated by PCR. Analysis of triplicate clones of each fragment, derived from 3 independent PCR's, has allowed the derivation of the entire nucleotide sequence of the BoNT/E gene. Translation of the sequence has shown BoNT/E to consist of 1252 amino acids, and as such represents the smallest BoNT characterised to date. The L chain of the toxin exhibits the highest level of sequence similarity to TeTx (40%). The L chains of BoNT/A and BoNT/D share 33% similarity with BoNT/E, while BoNT/C exhibits 32% similarity. In contrast, the TeTx H chain exhibits the lowest degree of homology (35%) with BoNT/E, with the BoNT H chains sharing 46%, 36% and 37%, for the type A, C and D neurotoxin types, respectively. Comparisons with partial amino acid sequences of the L chain of BoNT/E from *C. botulinum* strain Beluga and that from the strains Mashike, Iwanai and Otaru, indicate single amino acid differences in each case. Alignment of all characterised neurotoxins sequences (BoNT/A, BoNT/C, BoNT/D, BoNT/E and TeTx) shows them to be composed of highly conserved amino acid domains interspersed with amino acid tracts exhibiting little overall similarity. The most divergent region corresponds to the extreme COOH-terminus of each toxin, which may reflect differences in specificity of binding to neurone acceptor sites.

1.2.2 Results and Discussion

Probing with type A neurotoxin gene subfragments

To identify specific restriction fragments encoding principally L or H chain we initially sought to exploit DNA homology between the previously cloned BoNT/A gene (Thompson et al., 1990) and the BoNT/E gene. Two restriction fragments were gel-purified from the BoNT/A gene. The first, a 389 bp *Hpa*I-*Xho*II fragment, encoded amino acids 216 through 346 of the BoNT/A L chain. The second, a 628 bp *Hae*III-*Hin*dIII fragment, coded for amino acids 526 through 736 of the H chain (Thompson et al., 1990). Both fragments were radiolabelled and used in Southern blot experiments, employing type E genomic DNA cleaved with various restriction enzymes. Under aqueous conditions, it was established that hybridisation between the two genes occurred at 50°C in the case of the L chain probe, and at 53°C in the case of the H chain probe. The relatively low value of these figures was indicative of a fairly low level of homology between the genes in the regions probed, and, furthermore, suggested that homology was greater in the H chain encoding region.

Further experiments, in which the genomic DNA hybridised had been cleaved with a combination of endonucleases, allowed the derivation of crude restriction maps of the regions of the type E genome homologous to the type A probes employed (data not shown). Inexplicably, the two sets of results obtained could not be merged into a single unifying restriction map. This anomaly in the derived data meant that the coding potential of any particular fragment, with regard to the BoNT/E gene, could not be confidently assigned. A *different* route to cloning was therefore adopted.

Cloning of the L chain encoding region by PCR

By reference to published amino acid sequences of the NH_2-terminus of the BoNT/E H and L chains (Sathyamoorthy and Dasgupta, 1985; Schmidt et al., 1985), two oligonucleotides were synthesised (primers LE1 and HE1, Table 1) which would allow amplification of essentially the entire L chain encoding region by polymerase chain reaction (PCR). The nucleotides in positions of codon degeneracy were chosen on the basis of those most commonly found in clostridial genes (Young et al., 1989). PCR was undertaken with LE1 and HE1 and type E chromosomal DNA, at various temperatures, in buffer containing Mg^{2+} at final concentrations of either, 1.5, 2.2 or 3.0 mM . Agarose gel electrophoresis of the reaction products indicated that no specific DNA fragment had been generated. Previous comparative alignment of the BoNT/A and TeTx L chains (Thompson et al., 1990) had indicated that very few amino acids were absolutely conserved. One notable exception was a centrally located histidine rich motif. Indeed a preliminary amino acid sequence of part of the BoNT/E L chain confirmed that this motif was also present in BoNT/E (Wernars and Notermans, 1990). Two

Table 1. *Synthesised oligonucleotide primers employed in PCR-amplification of Clostridium botulinum NCTC 11219 genomic DNA.*

OLIGO	ABILITY TO PRIME	NUCLEOTIDE SEQUENCE [a]	NUCLEOTIDE POSITION [b] IN BoNT/E GENE	AMINO ACID [c] POSITION	REFERENCE
LE1	NO	I N S F N Y N D P 5'-ATAAATAGTTTTAATTATAAAGATCC-3' T T	237-262	BoNT/E, 4-12	Sathyamoorthy et al. (1985)
LE2	YES	H E L I H S L H G 5'-CACGAACTTATACATTCTCTACATGG-3' T T A AT	861-886	BoNT/E, 212-220	Wernars & Notermans (1990)
LE2'	YES	H E L I H S L H G 3'-GTGCTTGAATATGTAAGAGATGTACC-5' A A T TA	886-861	BoNT/E, 212-220	"
LE3	YES	F N Y N D P V N D 5'-TTTAATTATAATGATCCTGTAAATGA-3' T	246-271	BoNT/E, 7-15	Sathyamoorthy et al. (1985)
HE1	YES	I C I E I N N G E 3'-TATACATATCTTTATTTATTACCTCT-5' G A	1525-1501	BoNT/E_H, 3-11	"
HE2	YES	P Y I G P A L N I 5'-CCATATATAGGACCAGCATTAAATAT-3' T TT T	2037-2062	BoNT/A, 635-643	Thompson et al. (1990)
HE3	YES	K R N E K W D E V 5'-AAAAGAAATGAAAAATGGGATGAAGT-3' G G C A	2244-2269	BoNT/A, 701-709	"
HE4	YES	N K A M I N I N K F 5'-AATAAAGCAATGATAAATATAAATAAATT-3' TC T AT G C GG	2481-2509	BoNT/A, 778-887	"
HE5	YES	N R W I F V T I T N 3'-TTATCTACCTATAAACATTGTTATTGTTT-5' TC A A A	3189-3217	BoNT/A, 1012-1021	"
HE6	NO	G T K F L I K K Y 3'-CCTTGTTTTAAATATTATTTTTTTAT-5' A C T G C CA	3597-3622	BoNT/A, 1157-1165	"
HE7	NO	W E F I P V D D G W 3'-ACCCTTAAATATGGTCATCTACTTCCAACC-5' T G AA A T TG A T	3945-3974	BoNT/A, 1272-1291	"
LE4	YES	C R Q T Y I G Q Y 5'-GTAGGCAAACTTATATTGGACAGTA-3'	1267-1291	BoNT/E, 347-355	this study
HE8	YES	I V S N W M T K 3'-TATCATAGCTTAACCTACTGATTT-5'	2280-2303	BoNT/E, 685-692	this study
LE5	YES	T P D N Q F H I 3'-TGAGGTCTATTAGTTAAGGTATAAC-5'	582-606	BoNT/E, 119-126	this study
LE6	YES	L I T N I R G T 5'-CTAATAACAAATATAAGAGGTAC-3'	945-967	BoNT/E, 240-247	this study
HE9	YES	K N F S I S F W V R 3'-TTTTAAAATCATAATCAAAGACCCATTC-5' A	2965-2992	BoNT/E, 913-921	this study
HE10	YES	D N N S G W K V 5'-ATAATAATTCAGGATGGAAAGTAT-3'	3061-3084	BoNT/E, 945-952	this study

[a] the oligonucleotide primers LE1-LE3 and HE1-HE7 are "guessomers", designed to prime/anneal to DNA sequence encoding the amino sequence illustrated above. These amino sequences were either derived by NH_2-terminal sequencing of purified BoNT/E light- and heavy-chain subunits, or from the BoNT/A sequence, previously determined by recombinant means [Thompson et al., 1990]. Where these primers differed from the actual DNA sequence of the BoNT/E gene is illustrated below the sequence. With the exception of LE9, all other primers are perfect primers, based on the determined BoNT/E gene sequence. Primer LE9 is based on the equivalent region of the BoNT/F gene, which differs from the BoNT/E gene in this region by one nucleotide (unpublished data).

[b] position in the BoNT/E gene to which the oligonucleotides are targeted. Numbers correspond to nucleotide positions in Fig.2.

[c] numbering corresponds to the position of the amino acid sequences illustrated above the oligonucleotide sequences in either BoNT/E or BoNT/A.

further primers (LE2 and LE2', Table 1) were therefore synthesised corresponding to the sense and anti-sense DNA strand capable of encoding the histidine rich motif of the BoNT/E L chain. Subsequent PCR, at an annealing temperature of 37°C, using the primer pairs LE1 + LE2', and LE2 + HE1, resulted in an amplified DNA fragment of the expected size only in the case of the latter pair. Furthermore, appreciable amounts of DNA were only generated at the highest Mg^{2+} concentrations employed. These data suggested that the failure of the initial PCR to amplify a specific DNA fragment was due to inefficient priming of LE1. An alternative primer was therefore synthesised (LE3, Table 1), and used in combination with HE1 in a further PCR assay. In this case a DNA fragment of the expected size, 1.3 kb, was evident, following subsequent agarose gel electrophoresis of the reaction products.

The amplified products of the LE3 + HE1 reaction were blunt-ended with T4 polymerase and cloned into the *Sma*I site of pMTL20. Restriction analysis of 6 resultant recombinant plasmids indicated the presence of a common restriction fragment. Confirmation that the amplified fragment encoded BoNT/E was obtained by plasmid sequencing a representative plasmid recombinant (designated pCBE1) with both universal and reverse primer. Translation of the derived DNA sequences resulted in an uninterrupted amino acid sequence, which in the case of that derived using universal primer exhibited 100% identity with a preliminary BoNT/E sequence (Wernars and Notermans, 1990), while the sequence derived using reverse primer had substantial homology to the COOH-terminus of the BoNT/A L chain. Having established that the amplified fragment encoded BoNT/E, *the entire* nucleotide sequence of the pCBE1 insert was determined, as described in MATERIALS AND METHODS.

Cloning of H chain encoding DNA by PCR

In parallel to the experiments described above, a number of oligonucleotides were synthesised with a view to amplifying DNA regions of the neurotoxin gene encoding parts of the H chain. In the absence of amino acid sequence data for the BoNT/E H chain, we reasoned that amino acid motifs common to BoNT/A and TeTx may also be present in BoNT/E. The synthesised oligonucleotides (Table 1) therefore corresponded to a sense or anti-sense DNA strand capable of encoding amino acid motifs found in BoNT/A which were highly conserved in TeTx (Thompson et al., 1990). Individual PCR's were undertaken with all possible combinations of the sense and anti-sense oligonucleotides, under the conditions successfully established with the L chain primers. The only pairs of primers found to generate DNA fragments of the expected size were HE2 + HE5, HE3 + HE5, and HE4 + HE5. As the fragment derived from the reaction involving HE2 + HE5 was the largest (c. 1.2 kb), this particular DNA product was cloned, following blunt-ending, into the *Sma*I site of pMTL20. Plasmid sequencing of the resultant recombinant, pCBE2, and translation of the nucleotide

sequence obtained established the presence of uninterrupted amino acid sequences exhibiting significant homology to the BoNT/A H chain. Thereafter, the complete nucleotide sequence of the insert of pCBE2 was determined (see MATERIALS AND METHODS).

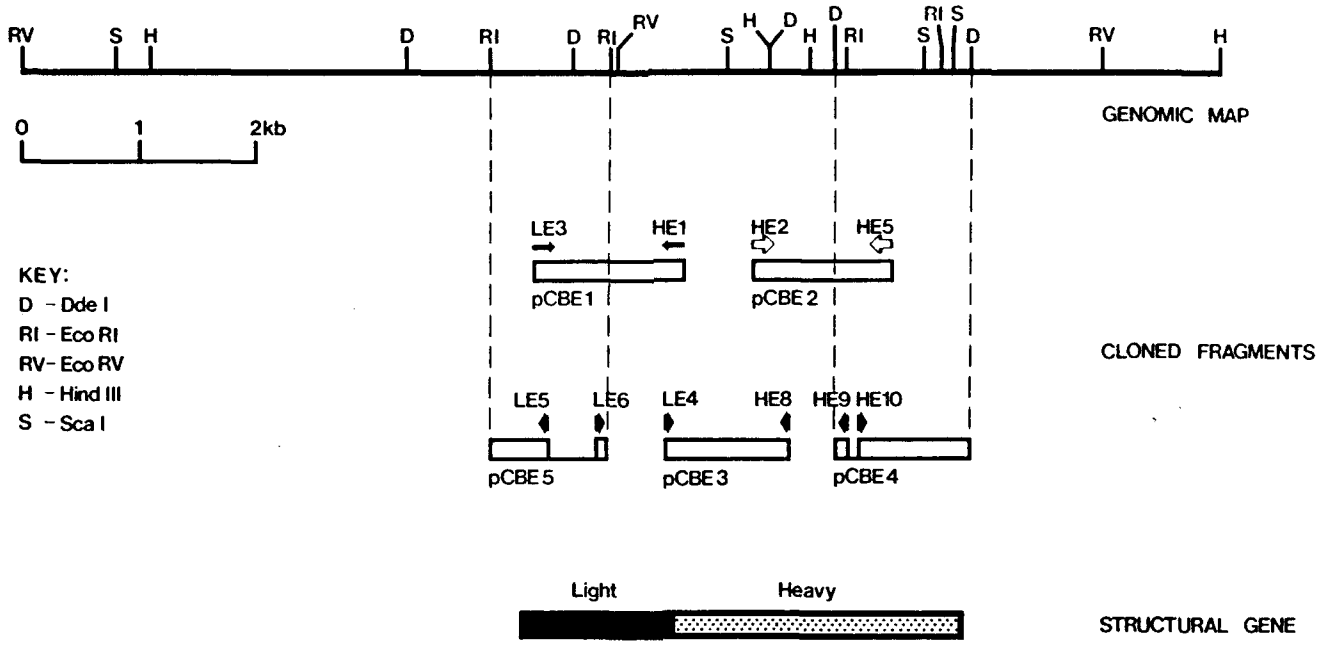


Fig. 1. *BoNT/E gene cloning strategy.* The 5 PCR-amplified regions of NCTC 11219 chromosome, that were cloned in the recombinant plasmids pCBE1-5, are represented by open boxes below the restriction map of the region of the genome encoding the BoNT/E gene. LE and HE primer sequences are given in Table 1. The arrows indicate the direction of DNA synthesis; solid arrows are perfect primers, open arrows guessomers. The vertical dotted line identifies the boundaries of the concatenated restriction fragment employed as the substrate for inverse PCR, using primer pairs LE5 + LE6 and HE9 + HE10.

Cloning of the remainder of the BoNT/E gene

To clone the intervening BoNT/E DNA between the inserts of pCBE1 and pCBE2, two oligonucleotides primers (LE4 and HE8; Table 1) were synthesised based on the determined nucleotide sequences of the pCBE1/2 inserts. The 1.03 kb product generated in a PCR using these primers was cloned directly into the specialised cloning vector pCR1000, and the nucleotide sequences of the inserts of a representative clone, pCBE3, determined.

DNA fragments carrying the remaining 3' and 5' ends of the BoNT/E gene were generated

by inverse PCR. This strategy required the identification of restriction sites proximal and distal to the gene. These sites were mapped by employing the radiolabelled PCR products generated by LE2 + HE1 and HE2 + HE5, in Southern blot experiments with restricted type E chromosomal DNA. The data obtained, together with information available from the nucleotide sequences of the inserts of pCBE1 and pCBE2, was used to construct an accurate restriction map of the region of the type E genome encoding the BoNT/E gene (Figure 1). This indicated that the 5' end of the structural gene resided on a c 1.0 kb *Eco*RI fragment, and that the 3' end of the gene was encompassed by a 1.1 kb *Dde*I fragment. Accordingly, type E chromosomal DNA was cleaved with the appropriate enzyme, self-ligated and a PCR undertaken on the circularised products using the oligonucleotide primer pairs LE5 + LE6 in the case of *Eco*RI cleaved DNA, and HE9 + HE10 in the case of DNA cut with *Dde*I. In both cases, DNA fragments of the calculated size were shown to be generated. Each amplified DNA product was cloned directly into pCR1000 (Mead et al., 1991), yielding pCBE4 (3'-end) and pCBE5 (5'-end), and the entire nucleotide sequences of their inserts determined (MATERIALS AND METHODS).

The complete nucleotide sequence of the BoNT/E gene

The 5 overlapping nucleotide sequences derived from the inserts of pCBE1 to pCBE5 in total encompassed the entire BoNT/E structural gene. However, because Taq polymerase is known to misincorporate nucleotides during DNA synthesis (Eckert and Kunkel, 1991), the sequence obtained may not have represented the authentic BoNT/E sequence. Therefore, all 5 cloned DNA fragments were reamplified by PCR, and cloned to give duplicate isolates of the five plasmids, pCBE1 to pCBE5. The nucleotide sequences of the entire inserts of each new plasmid were determined and compared to that derived from the initial clones. In those cases where a discrepancy in sequence was apparent, the appropriate fragment was PCR-amplified and cloned to give a third pCBE clone. The relevant region of the insert of this plasmid was then determined, and the consensus of the 3 sequences taken as being the correct BoNT/E gene sequence. The number of discrepancies in the three sequences was surprisingly high, with a total of 7 PCR-induced substitutions and 2 single base additions. Both of the latter, occurred in regions of the sequence composed of at least 5 consecutive 'A' nucleotides. This error rate equates to 7.8×10^{-4} per nt (ie., 9 errors per 11500 bases) and is most probably a direct result of the relatively high level of Mg^{2+} employed (Eckert and Kunkel, 1991).

The final sequence derived is illustrated in Fig. 2. The BoNT/E gene has a 75% A+T content and is composed of 1253 codons, initiating at nucleotide position 228 with a AUG codon and terminating at position 3986 with a UAA stop codon. The use of these particular translational initiation and termination signals is a general characteristic of clostridial genes

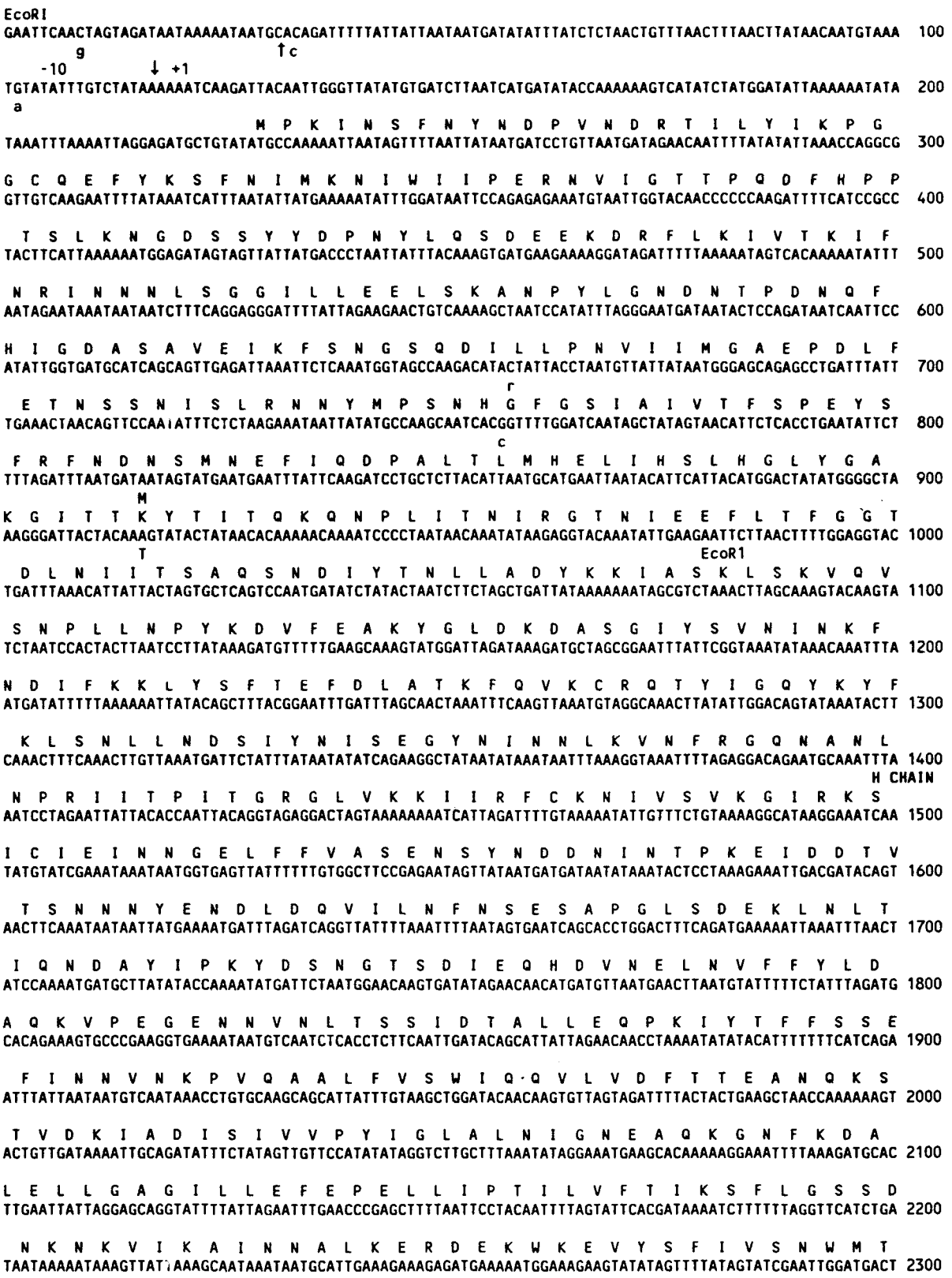
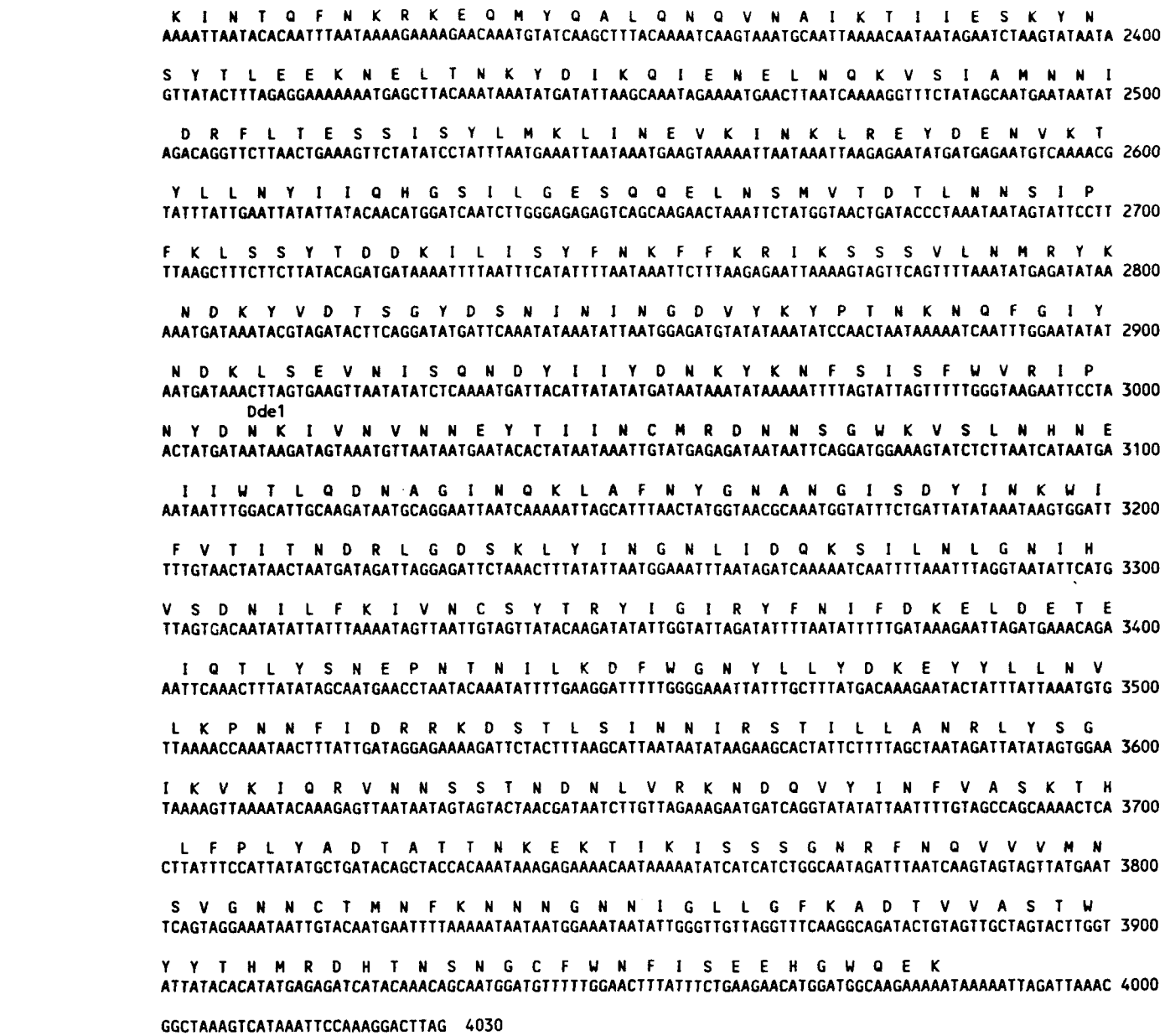


Fig 2. *Complete nucleotide sequence of the type E gene.* The BoNT/E amino acid sequence is given in the single letter code above the central nucleotide of the corresponding codon. Differences between the NCTC 11219 sequence and the partial nucleotide sequences of the genes of strain Beluga and



strains Mashike, Iwanai and Otaru, are indicated below the appropriate position of the sequence, in lower and upper case letters, respectively. An upward facing arrow indicates an insertion. Any change in the encoded amino acid is indicated above the NCTC BoNT/E amino acid sequence. The putative -10 promoter region (based on homology to the BoNT/A gene 5' non-coding region) and transcriptional initiation site are marked by a dashed line above the sequence and downward facing arrow, respectively. The ribosome binding site is indicated by a line above and below the sequence.

(Young et al., 1989). The AUG codon is preceded by a sequence typical of clostridial ribosome binding sites, in both its composition and distance (8 bases) from the AUG initiation codon. The codon usage exhibited by the gene is also typical of clostridial genes, with an extreme bias for codons ending in A and T, and the frequent use of codons recognised as

modulators of translation in *E. coli*. Although a number of sequences 5' to the BoNT/E structural gene exhibit some similarity to procaryotic promoter elements, assignment of such sequences as transcriptional signals will require appropriate experimental data. A reasonably high degree of sequence similarity (77.4% identity) does, however, exist between the 5' non-coding region of the BoNT/A (Binz et al., 1990) and BoNT/E gene. Based on this homology, the transcriptional start point would be nucleotide 117, and the TATATT motif at position 103 to 108 the putative -10 promoter element (Figure 2).

The sequence of a 983 bp portion of the BoNT/E gene (equivalent to nucleotides 1 to 988 of Fig. 2) , encoding part of the L chain, from a number of other *C. botulinum* type E strains has been reported, namely strain Beluga (Binz et al., 1990) and strains Mashike, Iwanai and Otaru (Fujii et al., 1990). The sequences derived from the latter 3 strains were identical and differ from that reported here for strain NCTC 11219 by a single nucleotide at position 916. Thus codon 230 of the BoNT/E genes from strains Mashike, Iwanai and Otaru is UAG, while in the BoNT/E gene of strain NCTC 11219, this codon is AAG. In contrast, the sequence derived from strain Beluga exhibits 4 nucleotide differences to the sequence of NCTC 11219. Three of these changes occur in the 5' non-coding region, including a single base 'C' insertion in the Beluga sequence (see Fig. 2), while the fourth difference results in a codon alteration of CGT (Beluga) to GGT (NCTC 11219) at position 756 (Fig. 2).

Comparative alignment of the nucleotide sequence of the two regions of the BoNT/A gene used as DNA probes in our original experiments, to the equivalent region of the BoNT/E gene, confirmed that a greater degree of DNA homology occurred between the H chain probe than the L chain probe. Thus, the 389 bp BoNT/A *Hpa*I-*Xho*II fragment exhibited 61.7% homology to the BoNT/E gene, whereas the 628 bp *Hae*III-*Hin*dIII fragment demonstrated 67.3% homology with the BoNT/E gene. The attainment of the complete nucleotide sequence of the BoNT/E gene also provided an opportunity to assess the reasons for the apparent ability/inability of the synthesised oligonucleotides to act as primers in PCR (Table 1). Such an assessment did not prove particularly informative. Thus, although the presence of 7 sequence mismatches in the case of HE6 may have precluded annealing to BoNT/E genomic DNA, 9 sequence mismatches in oligonucleotide HE4 apparently did not effect its ability to prime in PCR, assuming the generated fragment was indeed the region targeted. The success of primer HE4 may have been due to the fact that the 4 mismatches at the 3' end of the oligonucleotide would all have resulted in neutral d(G·T) pairing. More difficult to explain was the inability of LE1 to act as a primer (only 2 mismatches).

The complete amino acid sequence of the BoNT/E gene

The deduced amino acid sequence of BoNT/E demonstrated that the neurotoxin is comprised of 1252 residues, making it the smallest neurotoxin yet characterised. The amino acids at positions 423 through 435 demonstrated perfect agreement with those determined experimentally by NH_2-terminal sequencing of the purified BoNT/E H chain (Sathyamoorthy and Dasgupta, 1985; Schmidt et al., 1985). A more extensive recent sequence had indicated a presence of a single unassigned amino acid ("X") at $BoNT/E_H$ positions 16 and 19 (Dasgupta and Datta, 1988). The sequence deduced here indicates that the first "X" equates to the dipeptide sequence Ala-Ser, while the second "X" is a Ser residue. Comparisons between the NCTC 11219 L chain and the partial amino acid sequences of the BoNT/E L chains of strain eluga and strains Mashike, Iwanai and Otaru, indicated a single amino acid difference in each case. Thus, the Gly residue at position 177 in the NCTC 11219 toxin has been replaced by Arg in Beluga BoNT/E, while the Lys amino acid at position 230 in the NCTC 11219 BoNT/E is Met in the equivalent position of the three Japanese strain-derived toxins.

1.3 CLONING/ SEQUENCING OF THE BoNT/B GENE

1.3.1 Summary

DNA fragments derived from the *Clostridium botulinum* type A neurotoxin (BoNT/A) gene (*botA*) were used in DNA/DNA hybridisation reactions to derive a restriction map of the region of the *C. botulinum* type B strain Danish chromosome encoding *botB*. As the one probe encoded part of the BoNT/A heavy (H) chain, and the other part of the light (L) chain, the position and orientation of *botB* relative to this map was established. The temperature at which hybridisation occurred indicated that a higher degree of DNA homology occurred between the two genes in the H chain encoding region. Using the derived restriction map data, a 2.1 kb *Bgl*II-*Xba*I fragment encoding the entire BoNT/B L chain and 108 amino acids of the H chain was cloned and characterised by nucleotide sequencing. A contiguous 1.8 kb *Xba*I fragment encoding a further 623 amino acids of the H chain was also cloned. The 3'-end of the gene was obtained by cloning a 1.6 kb fragment amplified from genomic DNA by inverse polymerase chain reaction. Translation of the nucleotide sequence derived from all three clones demonstrated that BoNT/B was composed of 1291 amino acids. Comparative alignment of its sequence with all currently characterised BoNT's (A, C, D, E) and tetanus (TeTx) showed that a wide variation in percentage homology occurred dependent on which component of the dichain was compared. Thus, the L chain of BoNT/B exhibits the greatest degree of

homology (50% identity) with the TeTx L chain, whereas its H chain is most homologous (48% identity) with the BoNT/A H chain. Overall, the 6 neurotoxins were shown to be composed of highly conserved amino acid domains interceded with amino acid tracts exhibiting little overall similarity. In total 68 amino acids, out of an average of 442, are absolutely conserved between L chains and 110, out of 845 amino acids, between H chains. Conservation of Trp residues (1 in the L chain, and 9 in the H chain) was particularly striking. The most divergent region corresponds to the extreme carboxyterminus of each toxin, which may reflect differences in specificity of binding to neurone acceptor sites.

1.3.2 Results and Discussion

Southern blot analysis of the botB *gene*

A 389 bp *Hpa*I-*Xho*II *botA* fragment, encoding amino acids 216 through 346 of the BoNT/A L chain, and a 628 bp *Hae*II-*Hin*dIII fragment, coding for amino acids 526 through 736 of the H chain (Thompson et al., 1990), were radiolabelled and used in DNA/DNA hybridisations with type B chromosomal DNA cleaved with various restriction enzymes. Reactions were performed in aqueous solution over a range of temperatures. "Weak" hybridisation between the two genes was found to occur at 53°C and 56°C with the L and H chain probes, respectively (data not shown). The strength of the signal observed, and the relatively low stringency required were indicative of a fairly low level of DNA homology between *botA* and *botB*. Furthermore, these results suggest that the L chain encoding regions of the two genes are less homologous than the H chain encoding region, at least in the areas probed. Having established the conditions at which hybridisation occurred, the type B genomic DNA was cleaved with various combinations of restriction endonucleases and the nylon membranes carrying the resultant fragments sequentially hybridised with the two probes. The data obtained allowed the derivation of a restriction map of the region of the type B genome encoding *botB*. Furthermore the use of the two probes enabled the assignment of both the position of *botB* and its relative orientation, with respect to the derived map (Fig. 3).

Cloning and sequencing of the botB *L chain.*

The restriction map derived by the Southern blot experiments (Fig. 3) indicated that a 2.1 kb *Bgl*II-*Xba*I fragment principally encoded the L chain of BoNT/B. To clone this DNA, and to minimise the risk of cloning contiguous BoNT/B encoding regions, the targeted fragment was purified by a two-stage gel isolation procedure. *C. botulinum* type B chromosomal DNA

was cleaved with *Xba*I and fragments of approximately 7.5 kb in size purified from agarose gels by electroelution. The isolated DNA was then subjected to digestion with *Bgl*II, DNA fragments of around 2.1 kb in size gel-purified, ligated to similarly cut pMTL32 vector DNA

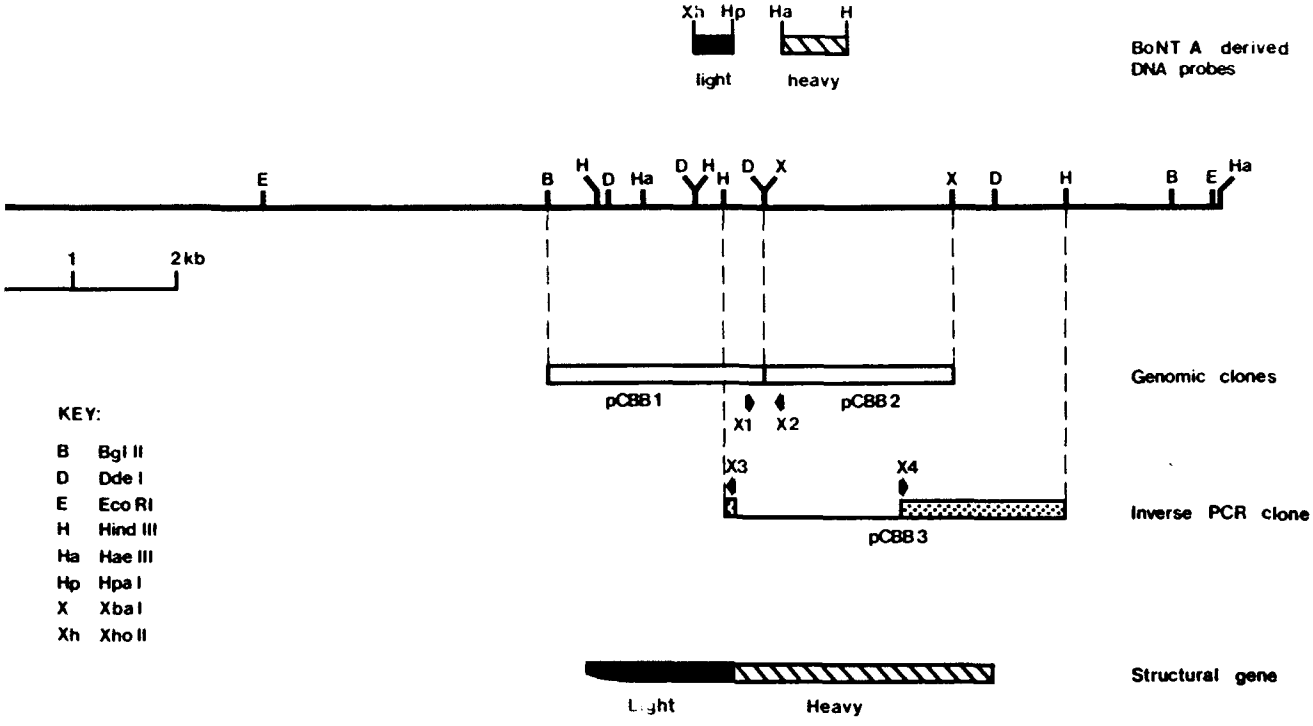


Fig. 3. ***Strategy employed in the cloning of the botB gene.*** The illustrated restriction map of the *C. botulinum* genome was generated using the indicated *botA* DNA fragments as probes in Southern blots. Regions of the strain B/Danish chromosome, that were cloned in the recombinant plasmids pCBB1 and pCBB2, are represented by open boxes below the restriction map. The cloned inserts of these plasmids were shown to be contiguous on the genome by PCR amplification of the region of the chromosome spanning their common *Xba*I site, using primers X1 (5'-CCAAGTGAAAATACAGAATCAC-3') and X2 (3'-CCCACTTTGTCTATCATTTA-5'), and sequencing across this junction. The insert of pCBB3 was derived by PCR amplification of *Hin*dIII cut, concatenated chromosomal DNA using primers X4 (5'-AT-AGAGATTTATATATTGGAG-3') and X3 (5'-TTATATACAGCCAAATGCTCCTTGC-3')

(Fig. 4), and the resultant TG1 transformants screened for the presence of recombinant clones using the *botA* L chain probe. The vector pMTL32 was specifically constructed for the purposes of cloning the *botB* DNA (see Fig. 4). Based on the pMTL1003 backbone (Brehm et al., 1992), it carries multiple cloning sites flanked on either side by tandem copies of transcriptional terminators. Heterologous genes inserted into the multiple cloning sites will therefore only be expressed if they carry indigenous transcriptional elements recognised by the RNA polymerases of *E. coli*.

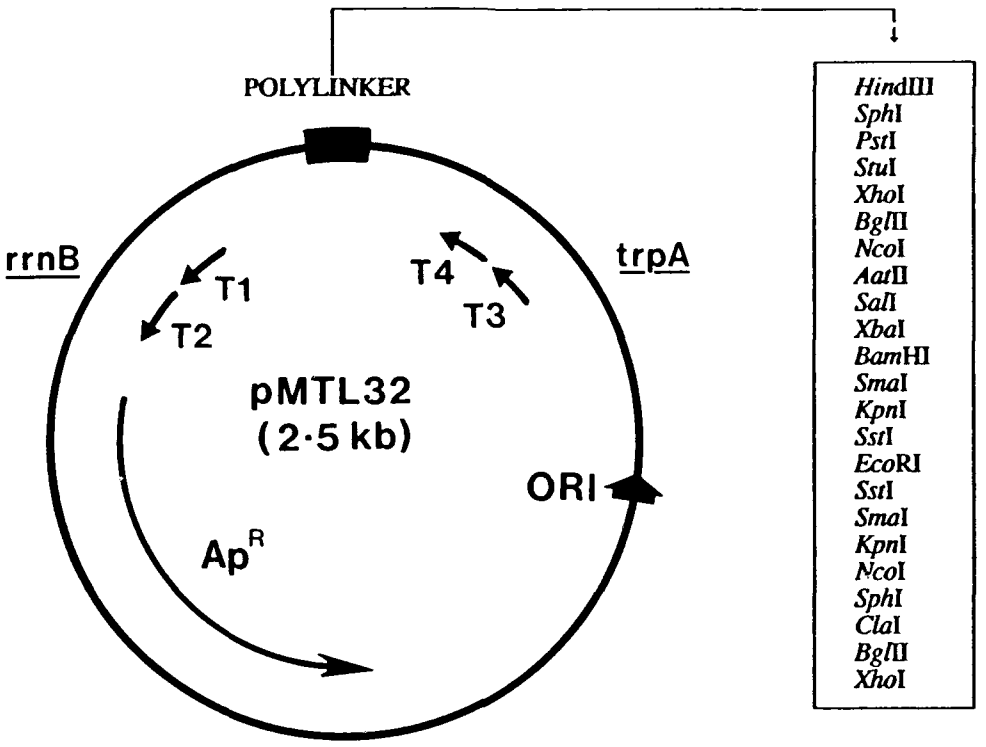


Fig. 4. The cloning vector pMTL32. This plasmid was derived as follows. A synthetic DNA fragment (5'-AGCCCGCCTAATGAGCGGGCTTTTTTT-3'), corresponding to the *E. coli trpA* transcriptional terminator, was ligated to *Stu*I -cleaved pMTL23 (Chambers et al., 1988) and a recombinant plasmid selected (pTRP23) in which two tandem copies of *trpA* had been inserted. The resultant double terminator, together with part of the pMTL23 polylinker region, was excised as a 107 bp *Nru*I-*Eco*RI fragment and inserted between the *Eco*RI and *Eco*RV sites of plasmid pMTL1003 (Brehm et al., 1991). As the c. 350 bp *Eco*RI-*Eco*RV fragment of pMTL1003 is deleted during this manipulation, the resultant plasmid, pMTL32, does not carry a copy of the *trp* promoter.

The recombinant plasmid obtained, designated pCBB1, was shown by digestion with appropriate endonucleases to contain restriction enzyme recognition sites consistent with the map illustrated in Fig. 3. Its entire insert was excised by digestion with *Bam*HI and *Bgl*II M13 recombinant templates containing random inserts derived using a sonication procedure (Minton et al., 1986). Using these templates, and custom synthesised oligonucleotides the entire nucleotide sequence of the insert was determined on both strands. Translation of the resultant sequence indicated the presence of an open reading frame (ORF) encoding a polypeptide of 549 amino acids in size. The aminoterminus of this polypeptide exhibited perfect conformity to that experimentally determined for purified BoNT/B L chain (Sathyamoorthy and DasGupta, 1985). Amino acids 442 through 459 were identical to that

determined for purified BoNT/B H chain (Sathymoorthy and DasGupta, 1985). Thus the insert carried by pCBB1 was deemed to encode the entire L chain of BoNT/B and 108 amino acids from the H chain.

Cloning and sequencing of the* botB *H chain.

Having established that the 2.1 *Bgl*II-*Xba*I fragment encoded the entire BoNT/B L chain and the aminoterminus of the H chain, it was apparent that the adjacent 1.8 kb *Xba*I fragment (Fig. 3) should encode the majority of the remaining H chain. Type B chromosomal DNA was cleaved with *Hin*dIII, fragments of approximately 3.5 kb isolated, digested with *Xba*I and fragments of around 1.8 kb in size gel purified. The isolated DNA was ligated with *Xba*I-cleaved pMTL32, transformed into *E. coli* TG1 and recombinant plasmids identified by probing with the radiolabelled *botA* H chain probe. One such plasmid was designated pCBB2, and the nucleotide sequence of its insert determined, following its insertion in M13mp18, by employing custom synthesised oligonucleotide primers.

Translation of the nucleotide sequence obtained revealed the presence of an continuous ORF of 623 codons, in the same reading frame relative to the *Xba*I site of that of the insert of plasmid pCBB1. To confirm that the two sequences were indeed contiguous a 289 bp region of DNA encompassing the *Xba*I site was amplified from type B genomic DNA using the primers X1 (5'-CCAAGTGAAAATACAGAATCAC-3') and X2 (3'-CCCACTTTGTCTATCATTTA-5') in a polymerase chain reaction (PCR), and cloned directly into ddT-tailed *Sma*I cut M13mp8. Nucleotide sequencing of a derivative template, using universal primer, demonstrated that the inserts of plasmids pCBB1 and pCBB2 were contiguous in the *C. botulinum* type B chromosome.

Completion of the* botB *sequence.

By combining the two sequences of pCBB1 and pCBB2, the derived contiguous ORF encoded 1170 amino acids, indicating that some 120 or so codons of the *botB* gene were missing. A DNA region encompassing the remaining 3'-end of the gene was cloned by inverse PCR. Type B chromosomal DNA was cleaved with *Hin*dIII, incubated with T4 ligase, and the resultant concatenated DNA used as a template in PCR with the oligonucleotides X3 (5'-ATAGAGATTTATATATTGGAG-3') and X4 (5'-TTATATACAGCCAAATGCTCCTTGC-3'). The 1.6 kb fragment generated was cloned directly into the specialised vector pCR1000 and the recombinant plasmid obtained designated pCBB3. A plasmid sequence reaction, undertaken with a primer previously employed in the determination of the nucleotide sequence

of the insert of plasmid pCBB2, confirmed the presence of the *botB* gene. Thereafter the nucleotide sequence of the region of pCBB3 encompassing the 3'-end of *botB* was determined by subcloning selected overlapping fragments into M13. To rule out the possibility that the insert of pCBB3 may have contained PCR-induced errors, a second version of this plasmid recombinant was derived by cloning the amplified DNA product from a further independent inverse PCR. Nucleotide sequencing of the appropriate regions of this second plasmid gave an identical sequence to that already derived from the primary isolate of pCBB3.

```
                                                RBS       M  P  V  T  I  N  N  F  N  Y  N  D  P  I  D  N
   1  AGCAATTTATGGCATTAAAAGGGATATAAACTTAAAATAAGGAGGAGAATATTTATGCCAGTTACAATAAATAATTTTAATTATAATGATCCTATTGATA

        N  N  I  I  M  M  E  P  P  F  A  R  G  T  G  R  Y  Y  K  A  F  K  I  T  D  R  I  W  I  I  P  E  R
 101  ATAATAATATTATTATGATGGAGCCTCCATTTGCGAGAGGTACGGGGAGATATTATAAAGCTTTTAAAATCACAGATCGTATTTGGATAATACCGGAAAG
                                                                  HindIII
        Y  T  F  G  Y  K  P  E  D  F  N  K  S  S  G  I  F  N  R  D  V  C  E  Y  Y  D  P  D  Y  L  N  T  N
 201  ATATACTTTTGGATATAAACCTGAGGATTTTAATAAAAGTTCCGGTATTTTTAATAGAGATGTTTGTGAATATTATGATCCAGATTACTTAAATACTAAT

        D  K  K  N  I  F  L  Q  T  M  I  K  L  F  N  R  I  K  S  K  P  L  G  E  K  L  L  E  M  I  I  N  G  I
 301  GATAAAAAGAATATATTTTTACAAACAATGATCAAGTTATTTAATAGAATCAAATCAAAACCATTGGGTGAAAAGTTATTAGAGATGATTATAAATGGTA

        P  Y  L  G  D  R  R  V  P  L  E  E  F  N  T  N  I  A  S  V  T  V  N  K  L  I  S  N  P  G  E  V  E
 401  TACCTTATCTTGGAGATAGACGTGTTCCACTCGAAGAGTTTAACACAAACATTGCTAGTGTAACTGTTAATAAATTAATCAGTAATCCAGGAGAAGTGGA

        R  K  K  G  I  F  A  N  L  I  I  F  G  P  G  P  V  L  N  E  N  E  T  I  D  I  G  I  Q  N  H  F  A
 501  GCGAAAAAAAGGTATTTTCGCAAATTTAATAATATTTGGACCTGGGCCAGTTTTAAATGAAAATGAGACTATAGATATAGGTATACAAAATCATTTTGCA

        S  R  E  G  F  G  G  I  M  Q  M  K  F  C  P  E  Y  V  S  V  F  N  N  V  Q  E  N  K  G  A  S  I  F  N
 601  TCAAGGGAAGGCTTCGGGGGTATAATGCAAATGAAGTTTTGCCCAGAATATGTAAGCGTATTTAATAATGTTCAAGAAAACAAAGGCGCAAGTATATTTA

        R  R  G  Y  F  S  D  P  A  L  I  L  M  H  E  L  I  H  V  L  H  G  L  Y  G  I  K  V  D  D  L  P  I
 701  ATAGACGTGGATATTTTTCAGATCCAGCCTTGATATTAATGCATGAACTTATACATGTTTTACATGGATTATATGGCATTAAAGTAGATGATTTACCAAT

        V  P  N  E  K  K  F  F  M  Q  S  T  D  A  I  Q  A  E  E  L  Y  T  F  G  G  Q  D  P  S  I  I  T  P
 801  TGTACCAAATGAAAAAAAATTTTTTATGCAATCTACAGATGCTATACAGGCAGAAGAACTATATACATTTGGAGGACAACATCCCAGCATCATAACTCCT

        S  T  D  K  S  I  Y  D  K  V  L  Q  N  F  R  G  I  V  D  R  L  N  K  V  L  V  C  I  S  D  P  N  I  N
 901  TCTACGGATAAAAGTATCTATGATAAAGTTTTGCAAAATTTTAGAGGGATAGTTGATAGACTTAACAAGGTTTTAGTTTGCATATCAGATCCTAACATTA

        I  N  I  Y  K  N  K  F  K  D  K  Y  K  F  V  E  D  S  E  G  K  Y  S  I  D  V  E  S  F  D  K  L  Y
1001  ATATTAATATATATAAAAATAAATTTAAAGATAAATATAAATTCGTTGAAGATTCTGAGGGAAAATATAGTATAGATGTAGAAAGTTTTGATAAATTATA

        K  S  L  M  F  G  F  T  E  T  N  I  A  E  N  Y  K  I  K  T  R  A  S  Y  F  S  D  S  L  P  P  V  K
1101  TAAAAGCTTAATGTTTGGTTTTACAGAAACTAATATAGCAGAAAATTATAAAATAAAAACTAGAGCTTCTTATTTTAGTGATTCCTTACCACCAGTAAAA
        HindIII
        I  K  N  L  L  D  N  E  I  Y  T  I  E  E  G  F  N  I  S  D  K  D  M  E  K  E  Y  R  G  Q  N  K  A  I
1201  ATAAAAAATTTATTAGATAATGAAATCTATACTATAGAGGAAGGGTTTAATATATCTGATAAAGATATGGAAAAAGAATATAGAGGTCAGAATAAAGCTA

        N  K  Q  A  Y  E  E  I  S  K  E  H  L  A  V  Y  K  I  Q  M  C  K  S  V  K  A  P  G  I  C  I  D  V
1301  TAAATAAACAAGCTTATGAAGAAATTAGCAAGGAGCATTTGGCTGTATATAAGATACAAATGTGTAAAAGTGTTAAAGCTCCAGGAATATGTATTGATGT
              HindIII
        D  N  E  D  L  F  F  I  A  D  K  N  S  F  S  D  D  L  S  K  N  E  R  I  E  Y  N  T  Q  S  N  Y  I
1401  TGATAATGAAGATTTGTTCTTTATAGCTGATAAAAATAGTTTTTCAGATGATTTATCTAAAAACGAAAGAATAGAATATAATACACAGAGTAATTATATA

        E  N  D  F  P  I  N  E  L  I  L  D  T  D  L  I  S  K  I  E  L  P  S  E  N  T  E  S  L  T  D  F  N  V
1501  GAAAATGACTTCCCTATAAATGAATTAATTTTAGATACTGATTTAATAAGTAAAATAGAATTACCAAGTGAAAATACAGAATCACTTACTGATTTTAATG

        D  V  P  V  Y  E  K  Q  P  A  I  K  K  I  F  T  D  E  N  T  I  F  Q  Y  L  Y  S  Q  T  F  P  L  D
1601  TAGATGTTCCAGTATATGAAAAACAACCCGCTATAAAAAAAATTTTTACAGATGAAAATACCATCTTTCAATATTTATACTCTCAGACATTTCCTCTAGA
                                                                                                  XbaI
        I  R  D  I  S  L  T  S  S  F  D  D  A  L  L  F  S  N  K  V  Y  S  F  F  S  M  D  Y  I  K  T  A  N
1701  TATAAGAGATATAAGTTTAACATCTTCATTTGATGATGCATTATTATTTTCTAACAAAGTTTATTCATTTTTTTCTATGGATTATATTAAAACTGCTAAT

        K  V  V  E  A  G  L  F  A  G  W  V  K  Q  I  V  N  D  F  V  I  E  A  N  K  S  N  T  M  D  K  I  A  D
1801  AAAGTGGTAGAAGCAGGATTATTTGCAGGTTGGGTGAAACAGATAGTAAATGATTTTGTAATCGAAGCTAATAAAAGCAATACTATGGATAAAATTGCAG

        I  S  L  I  V  P  Y  I  G  L  A  L  N  V  G  N  E  T  A  K  G  N  F  E  N  A  F  E  I  A  G  A  S
1901  ATATATCTCTAATTGTTCCTTATATAGGATTAGCTTTAAATGTAGGAAATGAAACAGCTAAAGGAAATTTTGAAAATGCTTTTGAGATTGCAGGAGCCAG
```

Fig 5. *Complete nucleotide sequence of the type B gene.* The illustrated sequence was derived by amalgamation of the derived nucleotide sequences of the inserts of pCBB1 to pCBB3 (Fig. 3). The

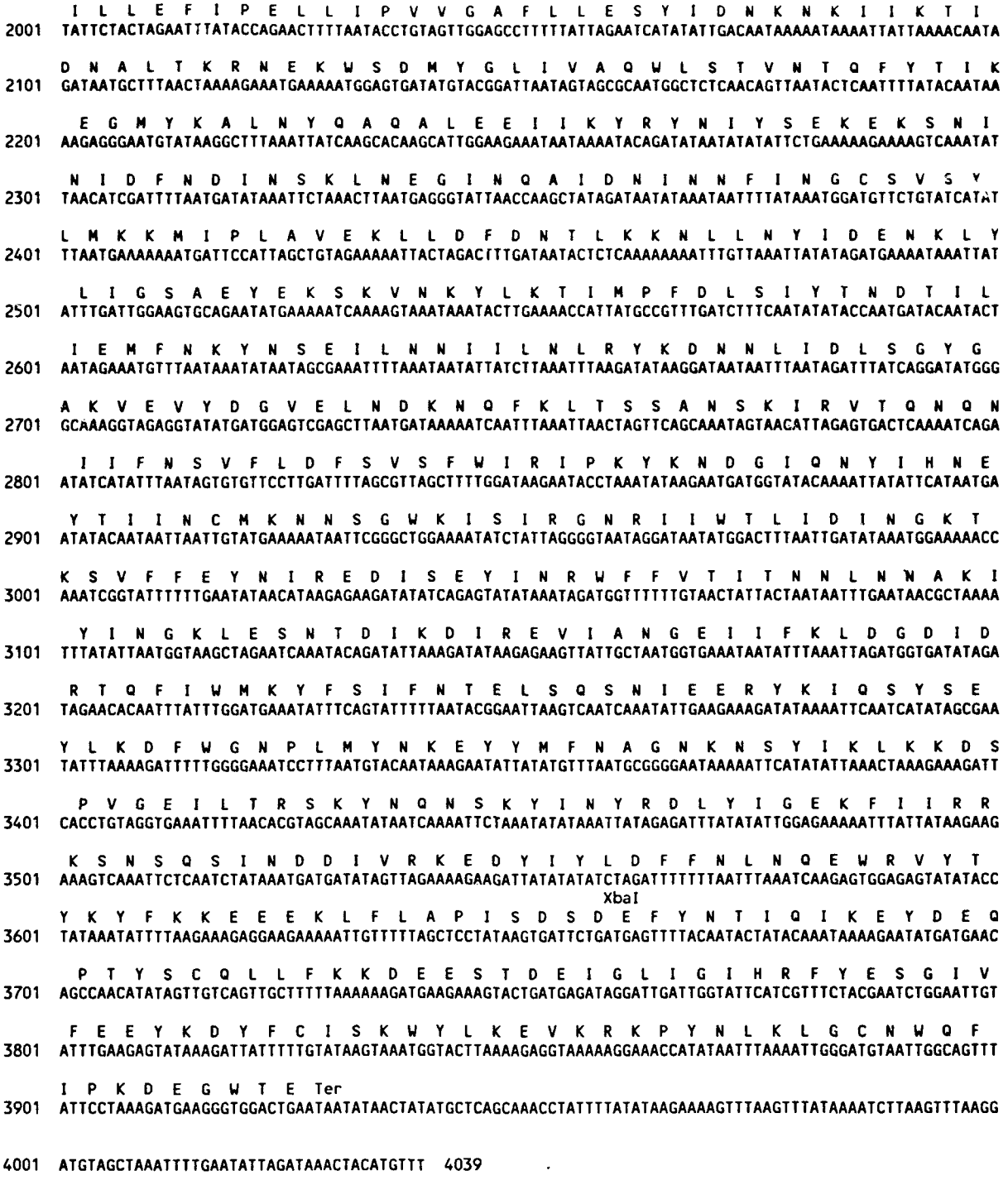


Fig 5. *Complete nucleotide sequence of the type B gene* (continued)
BoNT/B amino acid sequence is given in the single letter code above the first nucleotide of the corresponding codon. The ribosome binding site is indicated by a line above and below the sequence.

The entire nucleotide sequence of the *botB* gene (Fig. 5) was obtained by splicing the individual sequence information derived from the inserts of pCBB1, pCBB2 and pCBB3 into a contiguous sequence. The gene is composed of 1291 codons, initiating with an AUG codon at position 55 and terminating with a UAA stop codon at position 3928 (Fig. 5). The choice of these particular translational codons is typical of clostridial genes (Young et al., 1989). As

with all other *bot* genes characterised to date, the high A+T content of the DNA (74.6%) results in an extreme bias towards the use of codons ending in A or T, and the frequent use of codons recognised as modulators in *E. coli*. The translational start codon is preceded by a sequence typical of clostridial ribosome binding sites (Young et al., 1989).

Alignment of the nucleotide sequences of the two *botA*-derived DNA probes used in Southern blot mapping with the equivalent regions of *botB*, confirmed that the greater degree of homology existed in the respective H chain encoding regions over those encoding L chain. Specifically, the 628 bp *Hae*III-*Hin*dIII *botA* fragment demonstrated 65% homology with *botB*, whereas the 389 bp *Hpa*I-*Xho*II *botA* fragment had 54.8% homology with *botB*. Comparative alignment demonstrated that, in general, the overall DNA homology between the H chain and L chain encoding regions of all sequenced neurotoxin genes reflected the level of amino acid sequence homology (Table 2), and averaged between 50 to 60% identity. One consequence of this relative dissimilarity between genes is that DNA probes specific to each toxin gene may be easily designed. However, although there is sufficient homology in certain regions to derive a generalised probe for the generic detection of neurotoxin genes, it has not proven possible to design a probe which hybridises to all *bot* genes and not to the TeTx gene (unpublished data).

The complete amino acid sequence of BoNT/B.

The deduced primary sequence of BoNT/B demonstrates that the toxin is composed of 1291 amino acid residues. By comparison to partial amino acid sequences derived from purified polypeptides from other *C. botulinum* type B strains, it is apparent that variations in toxin structure occur. Thus although amino acid residues 2 through 17 exhibit perfect conformity to the sequence derived by Edman degradation of purified BoNT/B L chain of strain B/Okra (Sathyamoorthy and DasGupta, 1985), the amino acid at position 23 of the H chain was determined (DasGupta and Datta, 1988) to be Arg rather than the Ser residue seen here (position 464, Fig. 4). Similarly, the BoNT/B of strain B/657 possesses a Met amino acid at position 30 of the L chain (DasGupta and Datta, 1987) compared to Thr in the case of BoNT/B of Danish and B/Okra. Variations in the primary amino acid sequence of other types of BoNT have been noted, eg., between BoNT/A of strain 62A (Binz et al., 1990) and strain NCTC 2916 (Thompson et al., 1990), and between BoNT/E of strains Beluga, Mashike, Iwanai, Otaru and NCTC 11219 (this study). In the case of BoNT/B, such variations go some way to explaining observed dissimilarity in the immunological properties of BoNT/B isolated from different strains (Hatheway et al., 1981; Notermans et al., 1984).

1.4 CLONING/ SEQUENCING OF THE BoNT/F GENE

1.4.1 Summary

The oligonucleotide primers HE2 and HE5, previously employed in the PCR-mediated amplification of a 1.2 kb region of the *botE* gene, were used to amplify an equivalent region from the *botF* gene of the genome of a proteolytic *C. botulinum* type F strain. This amplified region was cloned into pMTL32 and, following the determination of its nucleotide sequence, used as a probe in Southern blot experiments to elucidate a restriction map of the type F genome encompassing the *botF* gene. The information was then used to assist in the cloning of 4 further overlapping fragments, amplified by PCR. Nucleotide sequence analysis of the inserts of the resultant plasmids (pCBF1-5) has allowed the derivation of the entire nucleotide sequence of the *botF* structural gene. Translation of the sequence revealed that BoNT/F is composed of 1278 amino acid residues. In relation to the other serotypes, the L chain exhibits the closest similarity to the L chains of BoNT/E (57%) and TeTx (43.5%), while the H chain most closely resembles BoNT/E (68%) and BoNT/A (44%). The nucleotide sequence of two other BoNT/F genes have also recently been determined, that of a non-proteolytic type F *C. botulinum* strain (ATCC 23387) and that of *C. baratii* ATCC 43756. All three toxins exhibit a surprising degree of divergence to each other. Thus, the L chain of the BoNT/F of strain Langeland shares 94% and 63% identity with ATCC 23387 and 43756, respectively. In contrast the H chains share 84% (ATCC 23387) and 79% (ATCC 43756) sequence identity.

1.4.2 Results and Discussion

Cloning of H chain encoding DNA by PCR

The oligonucleotide primers HE2 and HE5 (Table 1) had previously been shown to effect the amplification of a 1.2 kb fragment in a PCR using both type B and E DNA as template. When these two primers were employed in PCR using type F chromosomal DNA, an identically sized fragment was generated. This fragment was blunt-ended by treatment with T4 DNA polymerase and, following its isolation from an agarose gel, inserted into the *Sma*I site of pMTL32. The entire insert, and specific subfragments, were excised from the recombinant plasmid (pCBF1, Fig 6) and subcloned into M13mp18 and M13mp19. Templates prepared from the various recombinant phages were then subjected to nucleotide sequence analysis using universal primer. In certain instances the sequence obtained with a particular

template was extended using a synthesised sequence specific oligonucleotide. Translation of the nucleotide sequence obtained revealed the presence of a continuous ORF exhibiting substantial homology (74.4%) to BoNT/E.

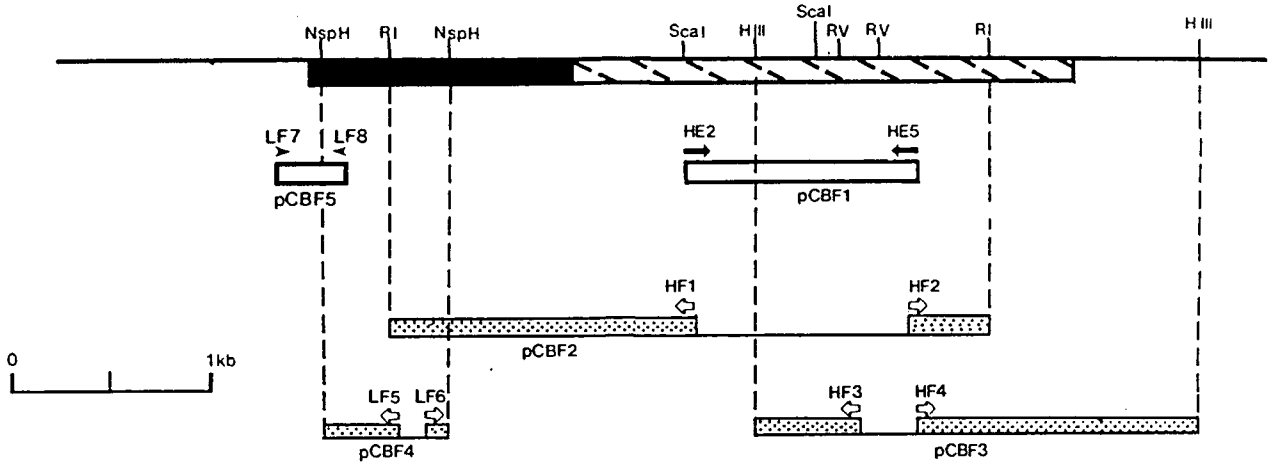


Fig. 6. *BoNT/F gene cloning strategy.* The 5 PCR-amplified regions of strain Langeland chromosome, that were cloned in the recombinant plasmids pCBF1-5, are represented by appropriate boxes below the restriction map of the region of the genome encoding the BoNT/F gene (bold line = light chain, hatched box = heavy chain). An open box (pCBF1 & pCBF5) indicates the amplified region was obtained in a standard PCR, the dotted boxes (pCBF2-4) represent regions amplified by inverse PCR. The vertical dotted lines identifies the boundaries of the concatenated restriction fragments employed as the substrate for inverse PCR, using primer pairs HF1 to HF4 and LF5 + LF6 (see text for sequences). Primers HE2 and HE5 are those used in the cloning of the equivalent region of *botE*. Abbreviated restriction sites are: RI, *Eco*RI; RV, *Eco*RV, and; HIII, *Hin*dIII.

Cloning of contiguous BoNT/F encoding DNA

To facilitate the cloning of regions of *botF* contiguous with that present in the insert of pCBF1, plasmid DNA was radiolabelled and used in Southern blot experiments to construct a restriction map of the type F genome (Fig. 6). The data obtained suggested that a 2.9 kb *Eco*RI fragment encompassed the cloned insert of pCBF1. As this fragment encoded significant further portions of the *botF* gene, it was targeted for cloning by a strategy involving inverse PCR. Type F chromosomal DNA was cleaved with *Eco*RI, incubated with T4 DNA ligase and the resultant concatenated DNA used as the template in a PCR with two oligonucleotides primers (HF1, 5'-CTCCTAATAATTCAAATGCCTCCTT-3'; HF2, 5'-AACTAGTTTTTAATTATACACAAAT-3') complementary to sequences at the proximal and

distal end of the pCBF1 insert (Fig. 6). The 1.9 kb fragment amplified was blunt-ended and cloned into the *Sma*I site of pMTL32 to yield the recombinant plasmid pCBF2. The insert of this plasmid was excised by digestion with *Bgl*II and M13 templates containing random inserts generated using the sonication procedure. The subsequent nucleotide sequence data obtained, in combination with that previously obtained from the pCBF1 insert, resulted in a contiguous sequence of 2,975 bp in length. Upon translation an uninterrupted ORF was evident encoding a polypeptide of 994 amino acid residues exhibiting 66.4% similarity to BoNT/E. From the alignment obtained between this polypeptide sequence and BoNT/E it was evident that a DNA region equivalent to some 150 codons was missing from the 3'-end of the *botF* gene, and approximately 122 codons from the 5'-end.

Cloning of the 5'- and 3'-end of botF

To identify restriction fragments encoding the 5'- and 3'-ends of *botF* Southern blot experiments were undertaken using type F genomic DNA cleaved with various restriction enzymes and two M13 recombinant clones, M13F16 and M13F44, as radiolabelled probes. These two M13 clones contained approximately 500 bp DNA inserts derived from either the proximal or distal ends of the sequenced 2.9 kb *Eco*RI fragment. Using probe M13F16 a *Nsp*HI fragment of approximately 650 bp in size was identified with the potential to encode the 5'-end of *botF*, while the probe M13F44 identified a 2.0 kb *Hin*dIII fragment deemed to carry the 3'-end of the gene. To clone the appropriate coding region of each fragment, PCR was undertaken with concatenated *Nsp*HI- and *Hin*dIII-cleaved type F chromosomal DNA with the primers LF5 (5'-TCAGGTCCTGCTCCCAATACAAGAAG-3') + LF6 (5'-CCCCGTTAGAAAACTAATGGATTCA-3') and HF3 (5'-TTACTACTATATATTCC-3') + HF4 (5'-GATCCAAGTATCTTAAAAGACTTTT-3'), respectively (Fig. 6). The fragments amplified (600 bp in the case of LF5 + LF6, and 1.5 kb in the case of HF3 + HF4) were cloned directly into pCR1000 to give the recombinant plasmids pCBF4 and pCBF3, respectively (Fig. 6).

The complete nucleotide sequence of the BoNT/F gene

The inserts of pCBF4, and a 0.8 kb *Eco*RI-*Hin*dIII subfragment of the pCBF3 insert, were subcloned into M13mp18 and M13mp19 and the resultant templates sequenced using a combination of universal primer and custom synthesised oligonucleotide primers. In the case of the pCBF3 -derived templates, the sequence obtained proved to be contiguous with that of the insert of pCBF2 and allowed the derivation of the missing 3'-end of the *botF* gene. In the

```
CAACTAGTAGATAACAAAAATAATGCAAAGAAGATGATAATTAGTAATAATATATTTATTTCCAATTGTTTAACTCTATCTTGTGGCGGTAAATATATAT  100

GTTTATCTATGAAAGATGAAAACTATAATTGGATGATATGTAATAATGAAAGCAACATACCTAAAAAGGCATATTTATGGACATTGAAAGAAGTATAGGG  200

       M  P  V  V  I  N  S  F  N  Y  N  D  P  V  N  D  D  T  I  L  Y  M  Q  I  P  Y  E  E  K  S  K
GGGATTTTATGCCAGTTGTAATAAATAGTTTTAATTATAATGACCCTGTTAATGATGATACAATTTTATACATGCAGATACCATATGAAGAAAAAAGTAA  300

 K  Y  Y  K  A  F  E  I  M  R  N  V  W  I  I  P  E  R  N  T  I  G  T  D  P  S  D  F  D  P  P  A  S
AAAATATTATAAAGCTTTTGAGATTATGCGTAATGTTTGGATAATTCCTGAGAGAAATACAATAGGAACGGATCCTAGTGATTTTGATCCACCGGCTTCA  400

 L  E  N  G  S  S  A  Y  Y  D  P  N  Y  L  T  T  D  A  E  K  D  R  Y  L  K  T  T  I  K  L  F  K  R  I
TTAGAGAACGGAAGCAGTGCTTATTATGATCCTAATTATTTAACCACTGATGCTGAAAAAGATAGATATTTAAAAACAACGATAAAATTATTTAAGAGAA  500

   N  S  N  P  A  G  E  V  L  L  Q  E  I  S  Y  A  K  P  Y  L  G  N  E  H  T  P  I  N  E  F  H  P  V
TTAATAGTAATCCTGCAGGGGAAGTTTTGTTACAAGAAATATCATATGCTAAACCATATTTAGGAAATGAACACACGCCAATTAATGAATTCCATCCAGT  600

 T  R  T  T  S  V  N  I  K  S  S  T  N  V  K  S  S  I  I  L  N  L  L  V  L  G  A  G  P  D  I  F  E
TACTAGAACTACAAGTGTTAATATAAAATCATCAACTAATGTTAAAAGTTCAATAATATTGAATCTTCTTGTATTGGGAGCAGGACCTGATATATTTGAA  700

N  S  S  Y  P  V  R  K  L  M  D  S  G  G  V  Y  D  P  S  N  D  G  F  G  S  I  N  I  V  T  F  S  P  E
AATTCTTCTTACCCCGTTAGAAAACTAATGGATTCAGGTGGAGTTTATGACCCAAGTAATGATGGTTTTGGATCAATTAATATCGTGACATTTTCACCTG  800

   Y  E  Y  T  F  N  D  I  S  G  G  Y  N  S  S  T  E  S  F  I  A  D  P  A  I  S  L  A  H  E  L  I  H
AATATGAATATACTTTTAATGATATTAGTGGAGGGTATAACAGTAGTACAGAATCATTTATTGCAGATCCTGCAATTTCACTAGCTCATGAATTGATACA  900

 A  L  H  G  L  Y  G  A  R  G  V  T  Y  K  E  T  I  K  V  K  Q  A  P  L  M  I  A  E  K  P  I  R  L
TGCACTGCATGGATTATACGGGGCTAGGGGAGTTACTTATAAAGAGACTATAAAAGTAAAGCAAGCACCTCTTATGATAGCCGAAAAACCCATAAGGCTA 1000

E  E  F  L  T  F  G  G  Q  D  L  N  I  I  T  S  A  M  K  E  K  I  Y  N  N  L  L  A  N  Y  E  K  I  A
GAAGAATTTTTAACCTTTGGAGGTCAGGATTTAAATATTATTACTAGTGCTATGAAGGAAAAAATATATAACAATCTTTTAGCTAACTATGAAAAAATAG 1100

  T  R  L  S  R  V  N  S  A  P  P  E  Y  D  I  N  E  Y  K  D  Y  F  Q  W  K  Y  G  L  D  K  N  A  D
CTACTAGACTTAGTAGAGTTAATAGTGCTCCTCCTGAATATGATATTAATGAATATAAAGATTATTTTCAATGGAAGTATGGGCTAGATAAAAATGCTGA 1200

 G  S  Y  T  V  N  E  N  K  F  N  E  I  Y  K  K  L  Y  S  F  T  E  I  D  L  A  N  K  F  K  V  K  C
TGGAAGTTATACTGTAAATGAAAATAAATTTAATGAAATTTATAAAAAATTATATAGCTTTACAGAGATTGACTTAGCAAATAAATTTAAAGTAAAATGT 1300

R  N  T  Y  F  I  K  Y  G  F  L  K  V  P  N  L  L  D  D  D  I  Y  T  V  S  E  G  F  N  I  G  N  L  A
AGAAATACTTATTTTATTAAATATGGATTTTTAAAAGTTCCAAATTTGTTAGATGATGATATTTATACTGTATCAGAGGGGTTTAATATAGGTAATTTAG 1400

  V  N  N  R  G  Q  N  I  K  L  N  P  K  I  I  D  S  I  P  D  K  G  L  V  E  K  I  V  K  F  C  K  S
CAGTAAACAATCGCGGACAAAATATAAAGTTAAATCCTAAAATTATTGATTCCATTCCAGATAAAGGTCTAGTGGAAAAGATCGTTAAATTTTGTAAGAG 1500

 V  I  P  R  K  G  T  K  A  P  P  R  L  C  I  R  V  N  N  R  E  L  F  F  V  A  S  E  S  S  Y  N  E
CGTTATTCCTAGAAAAGGTACAAAGGCGCCACCGCGACTATGCATTAGAGTAAATAATAGGGAGTTATTTTTTGTAGCTTCAGAAAGTAGCTATAATGAA 1600

N  D  I  N  T  P  K  E  I  D  D  T  T  N  L  N  N  N  Y  R  N  N  L  D  E  V  I  L  D  Y  N  S  E  T
AATGATATTAATACACCTAAAGAAATTGACGATACAACAAATCTAAATAATAATTATAGAAATAATTTAGATGAAGTTATTTTAGATTATAATAGTGAGA 1700

  I  P  Q  I  S  N  Q  T  L  N  T  L  V  Q  D  D  S  Y  V  P  R  Y  D  S  N  G  T  S  E  I  E  E  H
CAATACCTCAAATATCAAATCAAACATTAAATACACTTGTACAAGACGATAGTTATGTGCCAAGATATGATTCTAATGGAACAAGTGAAATAGAGGAACA 1800

 N  V  V  D  L  N  V  F  F  Y  L  H  A  Q  K  V  P  E  G  E  T  N  I  S  L  T  S  S  I  D  T  A  L
TAATGTTGTTGACCTTAATGTATTTTTCTATTTACATGCACAAAAAGTACCAGAAGGTGAAACTAATATAAGTTTAACTTCTTCAATTGATACGGCATTA 1900

S  E  E  S  Q  V  Y  T  F  F  S  S  E  F  I  N  T  I  N  K  P  V  H  A  A  L  F  I  S  W  I  N  Q  V
TCAGAAGAATCGCAAGTATATACATTCTTTTCTTCAGAGTTTATTAATACTATCAATAAACCTGTACACGCAGCACTATTTATAAGTTGGATAAATCAAG 2000

  I  R  D  F  T  T  E  A  T  Q  K  S  T  F  D  K  I  A  D  I  S  L  V  V  P  Y  V  G  L  A  L  N  I
TAATAAGAGATTTTACTACTGAAGCTACACAAAAAAGTACTTTTGATAAGATTGCAGACATATCTTTAGTTGTACCATATGTAGGTCTTGCTTTAAATAT 2100

 G  N  E  V  Q  K  E  N  F  K  E  A  F  E  L  L  G  A  G  I  L  L  E  F  V  P  E  L  L  I  P  T  I
AGGTAATGAGGTACAAAAAGAAAATTTTAAGGAGGCATTTGAATTATTAGGAGCGGGTATTTTATTAGAATTTGTGCCAGAGCTTTTAATTCCTACAATT 2200

L  V  F  T  I  K  S  F  I  G  S  S  E  N  K  N  K  I  I  K  A  I  N  N  S  L  M  E  R  E  T  K  W  K
TTAGTGTTTACAATAAAATCCTTTATAGGTTCATCTGAGAATAAAAATAAAATCATTAAAGCAATAAATAATTCATTAATGGAAAGAGAAACAAAGTGGA 2300

  E  I  Y  S  W  I  V  S  N  W  L  T  R  I  N  T  Q  F  N  K  R  K  E  Q  M  Y  Q  A  L  Q  N  Q  V
AAGAAATATATAGTTGGATAGTATCAAATTGGCTTACTAGAATTAATACACAATTTAATAAAAGAAAAGAACAAATGTATCAAGCTTTGCAAAATCAAGT 2400

 D  A  I  K  T  V  I  E  Y  K  Y  N  N  Y  T  S  D  E  R  N  R  L  E  S  E  Y  N  I  N  N  I  R  E
AGATGCAATAAAAACAGTAATAGAATATAAATATAATAATTATACTTCAGATGAGAGAAATAGACTTGAATCTGAATATAATATCAATAATATAAGAGAA 2500
```

```
E L N K K V S L A M E N I E R F I T E S S I F Y L M K L I N E A K V
GAATTGAACAAAAAAGTTTCTTTAGCAATGGAAAATATAGAGAGATTTATAACAGAGAGTTCTATATTTTATTTAATGAAGT(AATAAATGAAGCCAAAG 2600

 S K L R E Y D E G V K E Y L L D Y I S E H R S I L G N S V Q E L N
TTAGTAAATTAAGAGAATATGATGAAGGCGTTAAGGAATATTTGCTAGACTATATTTCAGAACATAGATCAATTTTAGGAAATAGTGTACAAGAATTAAA 2700

 D L V T S T L N N S I P F E L S S Y T N D K I L I L Y F N K L Y K
TGATTTAGTGACTAGTACTCTGAATAATAGTATTCCATTTGAACTTTCTTCATATACTAATGATAAAATTCTAATTTTATATTTTAATAAATTATATAAA 2800

K I K D N S I L D M R Y E N N K F I D I S G Y G S N I S I N G D V Y
AAAATTAAAGATAACTCTATTTTAGATATGCGATATGAAAATAATAAATTTATAGATATCTCTGGATATGGTTCAAATATAAGCATTAATGGAGATGTAT 2900

 I Y S T N R N Q F G I Y S S K P S E V N I A Q N N D I I Y N G R Y
ATATTTATTCAACAAATAGAAATCAATTTGGAATATATAGTAGTAAGCCTAGTGAAGTTAATATAGCTCAAAATAATGATATTATATACAATGGTAGATA 3000

 Q N F S I S F W V R I P K Y F N K V N L N N E Y T I I D C I R N N
TCAAAATTTTAGTATTAGTTTCTGGGTAAGGATTCCTAAATACTTCAATAAAGTGAATCTTAATAATGAATATACTATAATAGATTGTATAAGGAATAAT 3100

N S G W K I S L N Y N K I I W T L Q D T A G N N Q K L V F N Y T Q M
AATTCAGGATGGAAAATATCACTTAATTATAATAAAATAATTTGGACTTTACAAGATACTGCTGGAAATAATCAAAAACTAGTTTTTAATTATACACAAA 3200

 I S I S D Y I N K W I F V T I T N N R L G N S R I Y I N G N L I D
TGATTAGTATATCTGATTATATAAATAAATGGATTTTTGTAACTATTACTAATAATAGATTAGGCAATTCTAGAATTTACATCAATGGAAATTTAATAGA 3300

 E K S I S N L G D I H V S D N I L F K I V G C N D T R Y . G I R Y
TGAAAAATCAATTTCGAATTTAGGTGATATTCATGTTAGTGATAATATATTATTTAAAATTGTTGGTTGTAATGATACAAGATATGTTGGTATAAGATAT 3400

F K V F D T E L G K T E I E T L Y S D E P D P S I L K D F W G N Y L
TTTAAAGTTTTTGATACGGAATTAGGTAAAACAGAAATTGAGACTTTATATAGTGATGAGCCAGATCCAAGTATCTTAAAAGACTTTTGGGGAAATTATT 3500

 L Y N K R Y Y L L N L L R T D K S I T Q N S N F L N I N Q Q R G V
TGTTATATAATAAAAGATATTATTTATTGAATTTACTAAGAACAGATAAGTCTATTACTCAGAATTCAAACTTTCTAAATATTAATCAACAAAGAGGTGT 3600

 Y Q K P N I F S N T R L Y T G V E V I I R K N G S T D I S N T D N
TTATCAGAAACCAAATATTTTTTCCAACACTAGATTATATACAGGAGTAGAAGTTATTATAAGAAAAAATGGATCTACAGATATATCTAATACAGATAAT 3700

F V R K N D L A Y I N V V D R D V E Y R L Y A D I S I A K P E K I I
TTTGTTAGAAAAAATGATCTGGCATATATTAATGTAGTAGATCGTGATGTAGAATATCGGCTATATGCTGATATATCAATTGCAAAACCAGAGAAAATAA 3800

 K L I R T S N S N N S L G Q I I V M D S I G N N C T M N F Q N N N
TAAAATTAATAAGAACATCTAATTCAAACAATAGCTTAGGTCAAATTATAGTTATGGATTCAATAGGAAATAATTGCACAATGAATTTTCAAAACAATAA 3900

 G G N I G L L G F H S N N L V A S S W Y Y N N I R K N T S S N G C
TGGGGGCAATATAGGATTACTAGGTTTTCATTCAAATAATTTGGTTGCTAGTAGTTGGTATTATAACAATATACGAAAAAATACTAGCAGTAATGGATGC 4000

F W S F I S K E H G W Q E N .
TTTTGGAGTTTTATTTCTAAAGAGCATGGATGGCAAGAAAACTAATATAATAATTCAAAAAATAGGTATTAAAATAGAGGTAATATATATTACCCTCTAT 4100

TTTGGAATAATTTTAATATATTATATGAAACATATATAAATTTAAAGATAATATTAAATCAAGACACAAATTCAAATTAGAAATATAAAATGAAGTAAAT 4200

GAAAAGTGTAAAAAGTCATTAAATAAATTCAAAGACAGCATCTATATTTAAAAATTAGCAGTAATTCAAAGAATAGCTGCTATAAAAACATCATTAGTAG 4300

CTAGATTATTAACTTTTTGAAAAAATAAAAATAAATTTTTAGAATTTATACAAGACGATTTTTTATGTTTGTTGTAAAGCTT 4382
```

Fig 7. *Nucleotide sequence of the BoNT/F gene.* The illustrated sequence was derived by amalgamating the nucleotide sequences of the inserts of plasmids pCBF1 to pCBF5 (Fig. 6). The BoNT/F amino acid sequence is given in the single code above the first nucleotide of the corresponding codon.

case of the pCBF4-derived sequence, however, an alignment of the translated encoded polypeptide with BoNT/E revealed that the extreme 5'-end of the gene had not been cloned. It was estimated that, assuming BoNT/F has an identical number of amino acid residues at its aminoterminus to BoNT/E, 20 codons were missing from the start of the gene.

To obtained the missing region of *botF* the high degree of DNA homology between *botE* and *botF* was exploited by synthesising a "sense" strand oligonucleotide primer (LF7, 5'-CAACTAGTAGATAACAAAAATAATGC-3') based on the 5' non-coding region of *botE*. Following synthesis of a second anti-sense oligonucleotide primer (LF8, 5'-TGAGGTCCTGCTCCCAATACAAGAAG-3') based on the sequence derived from pCBF4, the missing region was amplified in a PCR and cloned into pCR1000 to give plasmid pCBF5 (Fig. 6). The complete sequence of the insert of pCBF5 was then determined by the plasmid sequencing procedure using universal and reverse primer. Having obtained the complete sequence of the *botF* gene, further representative clones of pCBF1 to pCBF5, or their equivalents, were derived from independent PCR's and their inserts sequenced. In those instances where a discrepancy arose the appropriate region of a third clone was examined. In total 12 discrepancies were noted, which represents a similar error rate to that seen during the cloning of the *botE* gene. The final sequence of the *botF* gene is illustrated in Fig.7. In nucleotide and codon composition, it is typical of the other characterised botulinum genes.

The complete amino acid sequence of BoNT/F

The deduced BoNT/F polypeptide is 1278 amino acid residues in length, putting it closer in size to BoNT/C (1276 aa) than any other neurotoxin. Although no amino acid sequence data has been derived for any BoNT/F toxin, the complete nucleotide sequences of the *botF* genes of a non-proteolytic strain of *C. botulinum* (ATCC 23387) and *C. baratii* (ATCC 43756) have recently been published (East et al., 1992; Thompson et al., 1993). Comparison of all three sequences reveal an unexpectedly high degree of divergence at both the nucleotide and amino acid sequence level (Table 2). The most divergent toxin is that of the *C. baratii* strain ATCC

	$F_{Langeland}$	$F_{ATCC\ 23387}$	$F_{ATCC\ 43756}$
$F_{Langeland}$	—	84	79
$F_{ATCC\ 23387}$	94	—	73
$F_{ATCC\ 43756}$	63	63	—

Table 2. *Amino acid homology between the L and H chain components of the 3 different types of BoNT/F.* Figures represent the % identity between di-chain components. The upper quadrant contains H chain comparisons, the lower L chain homologies. Langeland is a proteolytic group 1 *C. botulinum*, ATCC 23387 is a group 2 *C. botulinum* strain, and ATCC 43756 is a strain of *C. baratii*.

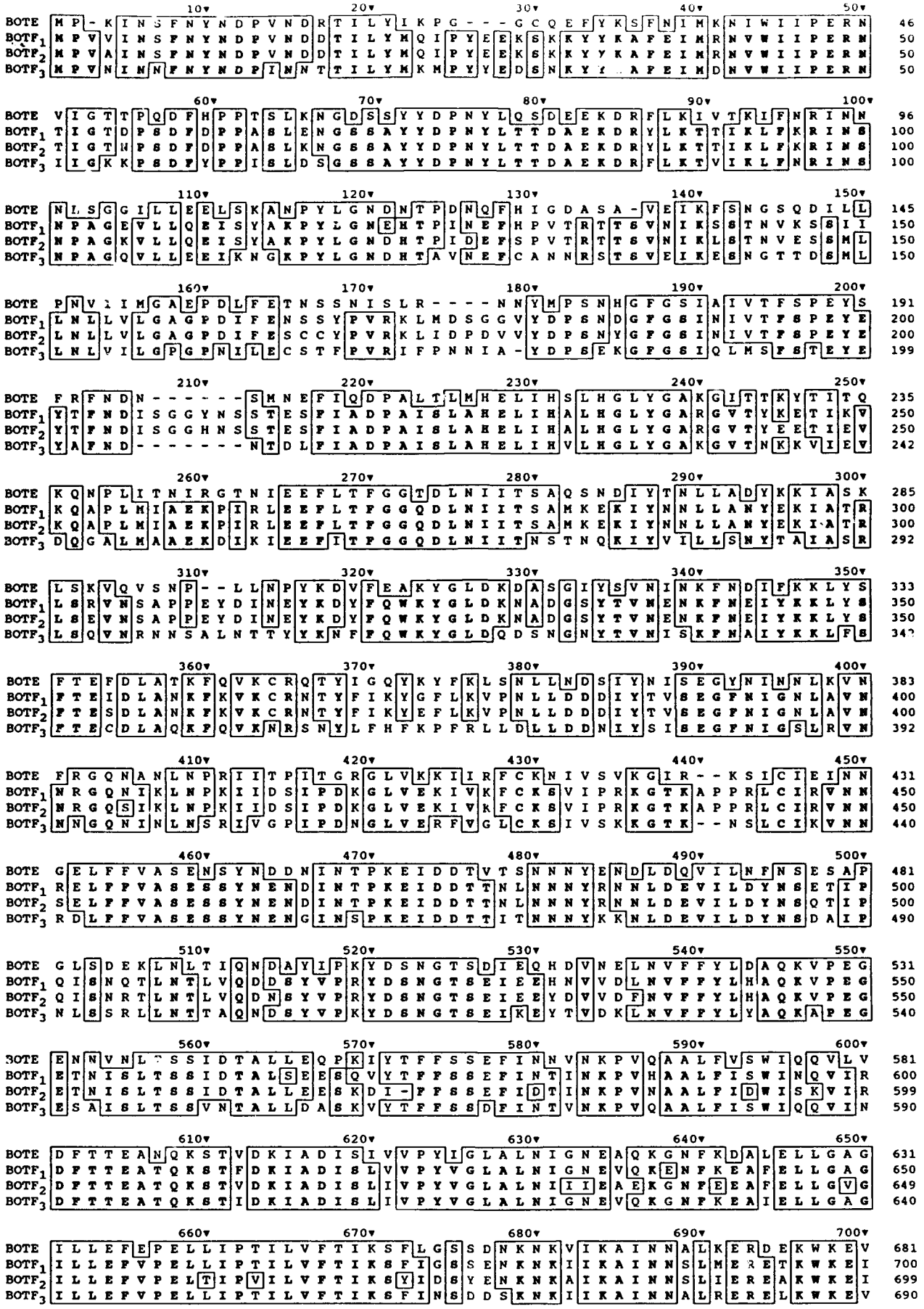


Fig. 8. *Amino acid sequence homology between the different BoNT/F's and BoNT/E*

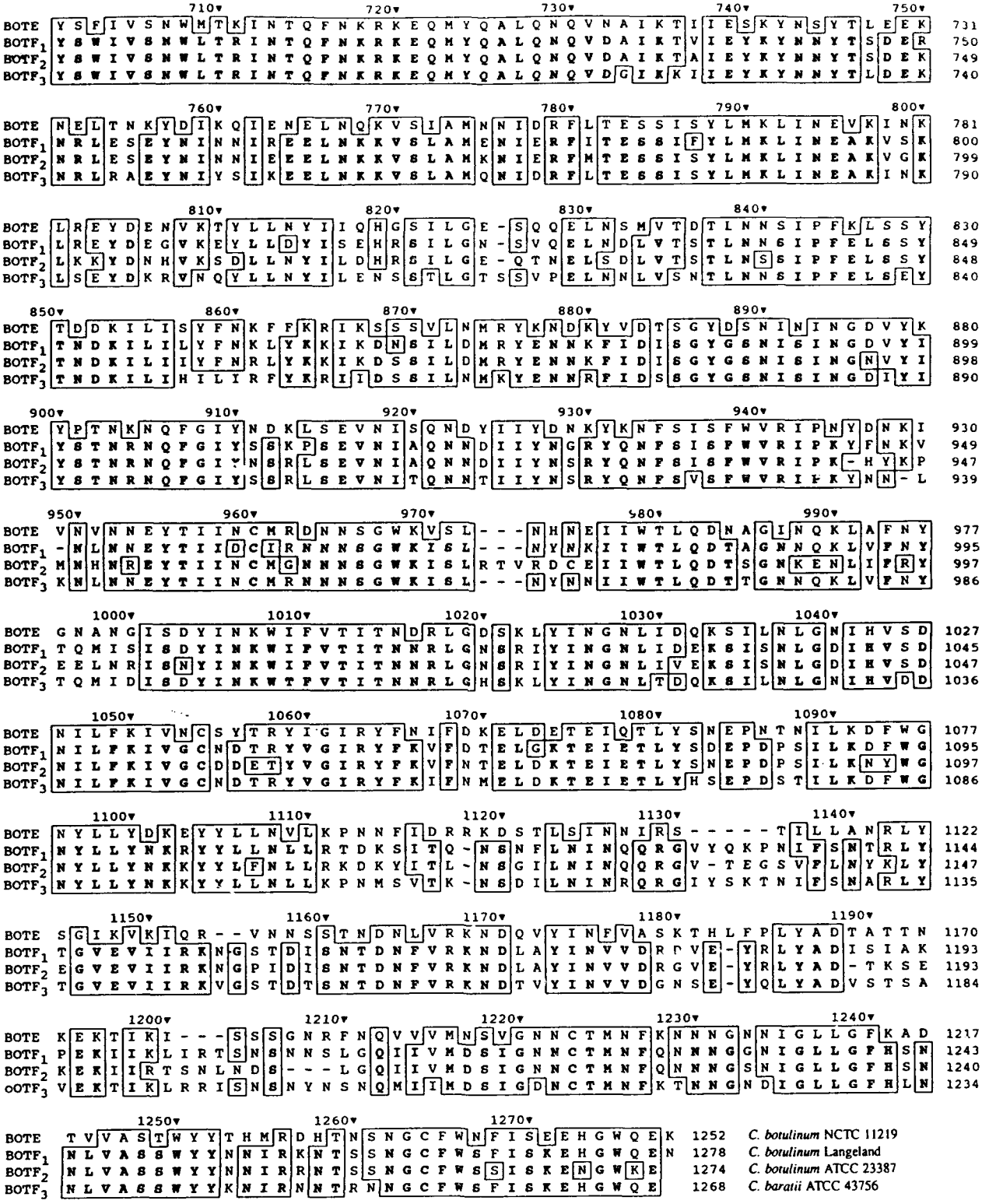


Fig. 8. ***Amino acid sequence homology between the different BoNT/F's and BoNT/E.*** The BoNT/E (BOTE) is that of the *C. botulinum* strain NCTC 11219. The BoNT/F's are: $BOTF_1$, the proteolytic *C. botulinum* strain Langeland; $BOTF_2$, the non-proteolytic strain ATCC 23387, and; $BOTF_3$ the *C. barati* strain ATCC 43756. Identical amino acids shared between at least 3 of the 4 toxins have been boxed. Those amino acids absolutely conserved between all 3 BoNT/F's are emboldened.

43756, which shares only 63% identity with the L-chain of both the proteolytic and non-proteolytic type F neurotoxins. Its H-chain is apparently more closely related to the former (79%) than the latter (73%). A complete alignment of all three type F neurotoxins, together with the related BoNT/E, is presented in Fig. 8. The DNA immediately 5' to the structural genes is conserved between all 3 organisms, eg., there are only 19 out of 273 mismatches between strain Langeland and ATCC 23387. In contrast, the regions immediately 3' to the structural genes appear completely unrelated. Most strikingly, sequence divergence begins immediately after the respective translational stop codons.

1.5 CLONING/ SEQUENCING OF THE BoNT/G GENE

1.5.1 Summary

The oligonucleotide primers LF7 and LE2, corresponding conserved motifs within the upstream 150 kDa non-toxic protein and the histidine-rich motif of all BoNT's, were used to amplify a 1.0 kb fragment encoding half of the BoNT/G L chain. This amplified region was cloned into pMTL20 and, following the determination of its nucleotide sequence, used as a probe in Southern blot experiments to determine a restriction map of the type G genome encompassing the *botG* gene. The information was then used to assist in the cloning of 8 further overlapping fragments, amplified by PCR. Nucleotide sequence analysis of the inserts of the resultant plasmids (pCBG1-9) has allowed the derivation of the entire nucleotide sequence of the *botG* structural gene. Translation of the sequence revealed that BoNT/G is composed of 1297 amino acid residues. In relation to the other serotypes, the neurotoxin is most closely related to BoNT/B. The L chains of these two toxins exhibit 61% identity. This is the highest degree of similarity seen between two neurotoxins of different serotypes. The observed similarity to BoNT/B continues into the H chain of BoNT/G, where the two toxins share 55% identity.

1.5.2 Results and Discussion

Cloning of the 5' end of an L chain encoding region of the BoNT/G gene

During the course of a parallel programme of work, in which oligonucleotide primers for the detection of toxin genes were being evaluated, it was noted that primers based on *botA* and

particularly *botB* sequence consistently amplified specific DNA fragments from the chromosomal DNA of the type G strain 89G. In one particular case the intensity and size of the fragment generated was equivalent to that seen with the intended target DNA, that of type A chromosome. This 1050 bp fragment was therefore cloned directly into pCR1000 and the proximal and distal regions of the insert of the resultant recombinant plasmid analysed in a plasmid sequence reaction using universal and reverse primer, respectively. A total of some 250 bp of sequence information was obtained with each primers, however, the two sequences proved to be identical to *botA*. It was concluded that the culture from which the chromosomal DNA had been prepared was contaminated with *C. botulinum* type A.

As an alternative, use was made of two previously synthesised primers, LF7 and LE2, employed during the cloning of other BoNT genes. The former was based on a conserved sequence motif found within the 5' non-coding region of *botE* and the *botF* gene of strain Langeland (see section 1.4). The latter corresponds to the histidine-rich motif of the L chain, and was employed during the cloning of the *botE* gene (see section 1.2). The use of these two primers in a PCR using 89G DNA as a template resulted in the amplification of a DNA fragment of the expected size, approx. 1.0 kb. This was cloned into the plasmid pCR1000 to give plasmid pCBG1 (Fig. 9). Subsequent nucleotide sequence analysis of the insert of pCBG1 confirmed that the amplified fragment was specific to the *botG* gene, encoding 237 amino acids from the NH_2-terminus of the BoNT/G L chain. Notably, the encoded polypeptide exhibited a high degree of sequence identity (58%) to the BoNT/B L chain.

Cloning of a contiguous region of the botG gene

Having cloned part of *botG*, experiments were undertaken to construct a restriction map of the region of strain 89G's genome carrying the gene. PCR was undertaken using LE2' (the complementary oligonucleotide of LE2) and a primer (HG1) based on a conserved motif (KDFWGN, position 1085-1090 in BoNT/B) found some 200 amino acids from the COOH-terminus of the H chain (Fig. 10). In view of the high degree of homology of the pCBG1 insert to the BoNT/B gene, the sequence of HG1 (5'-ATTTCCCCAAAAATCTTTTA-3)' was based on *botB*. The expected 2.65 kb DNA fragment amplified was used in addition to the insert of pCBG1 as a radiolabelled probe in Southern blots against restricted 89G chromosome. The use of two probes allowed the neurotoxin gene to be orientated relative to the restriction map obtained (Fig. 9). From this data it was apparent that the amplified 2.65 kb fragment could be cleaved approximately in half by the action of the endonuclease *Sca*I. Accordingly, the DNA sample obtained from a PCR using HG1 and LE2' was restricted with *Sca*I and the two DNA fragments generated gel purified and cloned into dT-tailed *Sma*I-cut pMTL21. The plasmid carrying the larger 1.5 kb

fragment was designated pCBG4, the plasmid carrying the smaller 1.2 kb fragment was designated pCBG3 (Fig. 9).

The nucleotide sequences of the inserts of PCBG3 and pCBG4 were derived by excising their inserts, transferring them to M13, and then "walking" along from each end using custom synthesised oligonucleotides as primers. To sequence across the *Sca*I site, a 100 bp fragment spanning this site was amplified from 89G DNA using appropriate oligonucleotides

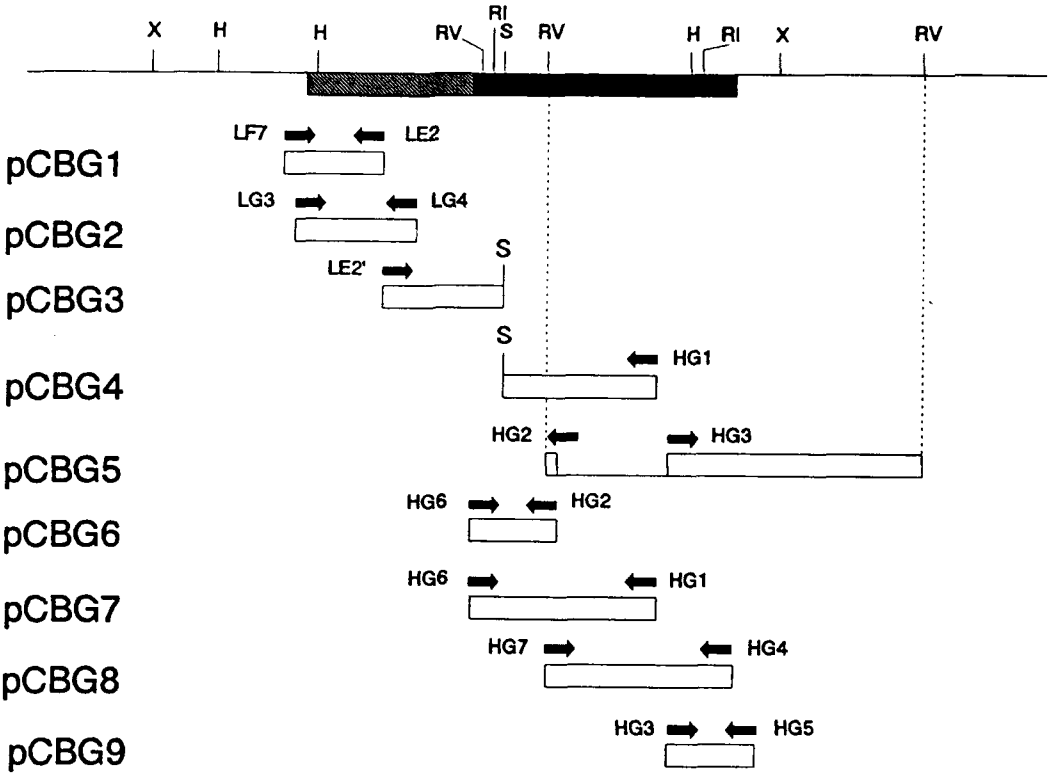


Fig. 9. *BoNT/G gene cloning strategy.* The 9 PCR-amplified regions of strain 89G chromosome, that were cloned in the recombinant plasmids pCBG1-9, are represented by open boxes below the restriction map of the region of the genome encoding the BoNT/G gene. LE primer sequences are given in Table 1. Primer LF7 was used during the cloning of the type F gene (section 1.4.2). All other primers are described in this section (1.5.2). The arrows indicate the direction of DNA synthesis. The vertical dotted line identifies the boundaries of the concatenated restriction fragment employed as the substrate for inverse PCR, using primer pairs LG2 + HG3. The fragment amplified in PCR using primers LE2' and HG1 was subsequently cleaved with *Sca*I and the resultant two DNA fragments cloned independently. Restriction enzyme sites are: h, *Hin*dIII; R1, *Eco*R1; RV, *Eco*RV; S, *Sca*I, and; X, *Xba*I.

and sequenced directly. In addition, a second clone, pCBG2 (Fig. 9), carrying 89G-derived DNA covering a similar region to that of the insert of pCBG1 was derived. In this case, however, the cloned fragment was amplified using oligonucleotides based on *botG* sequence

obtained from the clones pCBG1 (LG4, 5'-TAGGATCATGTCCTCCGAATG-3') and pCBG3 (LG3, 5'-CTATTTGGTATGCTATTTGTG-3'). The positioning of primer LG4 allowed the derivation of the authentic sequence of the histidine-rich motif which in clone pCBG1 equated to primer LE2.

Cloning of the 3'-end of the botG gene

To clone the missing 3'-end of the gene two "outward-facing" primers were synthesised (HG2, 5'-CGTTGAGAGCCACTGCGATAC-3'; HG3, 5'-GGTAGAGAATTAAATGCTAC -AG-3'), based on data obtained from pCBG4, and used in an inverse PCR with concatenated, *Eco*RV-cleaved, 89G chromosomal DNA. The 2.4 kb fragment generated was cloned into pCRII to give plasmid pCBG5 (Fig. 9). Although nucleotide sequencing of the clone obtained provided the sequence of the extreme 3'-end of *botG*, the sequence of the complete gene was not obtained as the clone contained a 400 bp deletion (nucleotides 3478 to 3894 on Fig. 10). Two further regions of the 89G genome were therefore amplified and cloned. Initially a 0.7 kb fragment was amplified, using HG3 and a primer (HG5, 5'-CCACACCTTTTATTTTA-3') based on the 3' non-coding region of the gene (determined using pCBG5), and cloned into pCRII to give pCBG9 (Fig. 9). Thereafter a second plasmid was similarly obtained, pCBG8 (Fig. 9), by cloning a 1.7 kb fragment amplified using two further primers, HG4 (5'-GGTATCCCAAACATATC-3') and HG7 (5'-ATGACGATATCCAATGC-3'). The insert of pCBG6 carries, as a contiguous region, parts of the inserts of pCBG4 and pCBG5 (Fig. 9).

Completion of the sequence

As as been the case with the previous *bot* genes cloned using PCR, data generated from a single clone cannot be relied upon due to the high incidence of PCR-induced errors. Further representative clones of each type were therefore obtained, by cloning appropriate fragments from independently performed PCRs, and their inserts characterised by sequencing. The data generated from these clones has been complied into a single sequence using DNASTAR software and is illustrated in Figure 10. In total, 7 nucleotide substitutions and a single nucleotide deletion were noted out of 15600 bp, an error rate of 4.5×10^{-4} per nt.

The deduced BoNT/G polypeptide is composed of 1291 amino acids, making BoNT/G one of the largest of the clostridial neurotoxins. Comparative analysis demonstrates that there is a remarkable degree of similarity between BoNT/G and BoNT/B, particularly between L chains where the percentage identity is 61%. This is the highest degree of homology seen between two immunologically distinct neurotoxins. This similarity is also shared by the respective

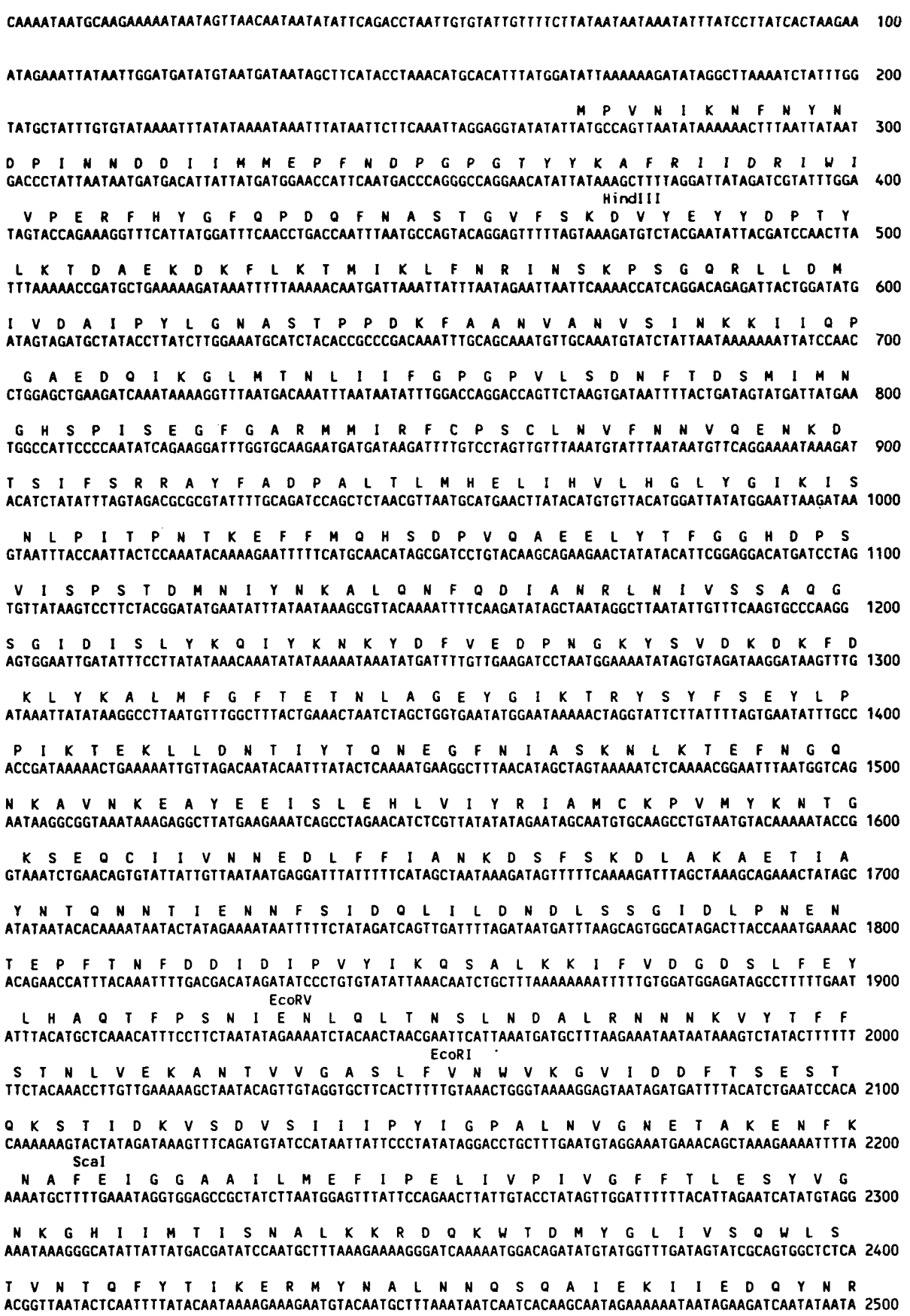


Fig 10. *Nucleotide sequence of the BoNT/G gene*

```
  Y  S  E  E  D  K  M  N  I  N  I  D  F  N  D  I  D  F  K  L  N  Q  S  I  N  L  A  I  N  N  I  D  D
GATATAGTGAAGAAGATAAAATGAATATTAACATTGATTTTAATGATATAGATTTTAAACTTAATCAAAGTATAAATTTAGCAATAAACAATATAGATGA 2600

  F  I  N  Q  C  S  I  S  Y  L  M  N  R  M  I  P  L  A  V  K  K  L  K  D  F  D  D  N  L  K  R  D  L
TTTTATAAACCAATGTTCTATATCATATCTAATGAATAGAATGATTCCATTAGCTGTAAAAAAGTTAAAAGACTTTGATGATAATCTTAAGAGAGATTTA 2700

  L  E  Y  I  D  T  N  E  L  Y  L  L  D  E  V  N  I  L  K  S  K  V  N  R  H  L  K  D  S  I  P  F  D  L
TTGGAGTATATAGATACAAATGAACTATATTTACTTGATGAAGTAAATATTCTAAAATCAAAAGTAAATAGACACCTAAAAGACAGTATACCATTTGATC 2800

  S  L  Y  T  K  D  T  I  L  I  Q  V  F  N  N  Y  I  S  N  I  S  S  N  A  I  L  S  L  S  Y  R  G  G
TTTCACTATATACCAAGGACACAATTTTAATACAAGTYTTTAATAATTATATTAGTAATATTAGTAGTAATGCTATTTTAAGTTTAAGTTATAGAGGTGG 2900

  R  L  I  D  S  S  G  Y  G  A  T  M  N  V  G  S  D  V  I  F  N  D  I  G  N  G  Q  F  K  L  N  N  S
GCGTTTAATAGATTCATCTGGATATGGTGCAACTATGAATGTAGGTTCAGATGTTATCTTTAATGATATAGGAAATGGTCAATTTAAATTAAATAATTCT 3000

  E  N  S  N  I  T  A  H  Q  S  K  F  V  V  Y  D  S  M  F  D  N  F  S  I  N  F  W  V  R  T  P  K  Y  N
GAAAATAGTAATATTACGGCACATCAAAGCAAATTCGTTGTATATGATAGTATGTTTGATAATTTTAGCATTAACTTTTGGGTAAGGACTCCTAAATATA 3100

  N  N  D  I  Q  T  Y  L  Q  N  E  Y  T  I  I  S  C  I  K  N  D  S  G  W  K  V  S  I  K  G  N  R  I
ATAATAATGATATACAAACTTATCTTCAAAATGAGTATACAATAATTAGTTGTATAAAAAATGACTCAGGATGGAAAGTATCTATTAAGGGAAATAGAAT 3200

  I  W  T  L  I  D  V  M  Q  N  L  N  Q  Y  F  S  N  I  G  I  K  D  N  I  S  D  Y  I  N  K  W  F  S
AATATGGACATTAATAGATGTAATGCAAAATCTAAATCAATATTTTTCGAATATAGGTATAAAAGATAATATATCAGATTATATAAATAAATGGTTTTCC 3300

  I  T  I  T  N  D  R  L  G  N  A  N  I  Y  I  N  G  S  L  K  K  S  E  K  I  L  N  L  D  R  I  N  S  S
ATAACTATTACTAATGATAGATTAGGTAACGCAAATATTTATATAAATGGAAGTTTGAAAAAAAGTGAAAAAATTTTAAACTTAGATAGAATTAATTCTA 3400

  N  D  I  D  F  K  L  I  N  C  T  D  T  T  K  F  V  W  I  K  D  F  N  I  F  G  R  E  L  N  A  T  E
GTAATGATATAGACTTCAAATTAATTAATTGTACAGATACTACTAAATTTGTTTGGATTAAGGATTTTAATATTTTTGGTAGAGAATTAAATGCTACAGA 3500

  V  S  S  L  Y  W  I  Q  S  S  T  N  T  L  K  D  F  W  G  N  P  L  R  Y  D  T  Q  Y  Y  L  F  N  Q
AGTATCTTCACTATATTGGATTCAATCATCTACAAATACTTTAAAAGATTTTTGGGGGAATCCTTTAAGATACGATACACAATACTATCTGTTTAATCAA 3600

  G  M  Q  N  I  Y  I  K  Y  F  S  K  A  S  M  G  E  T  A  P  R  T  N  F  N  N  A  A  I  N  Y  Q  N  L
GGTATGCAAAATATCTATATAAAGTATTTTAGTAAAGCTTCTATGGGGGAAACTGCACCACGTACAAACTTTAATAATGCAGCAATAAATTATCAAAATT 3700
                                  HindIII
  Y  L  G  L  R  F  I  I  K  K  A  S  N  S  R  N  I  N  N  D  N  I  V  R  E  G  D  Y  I  Y  L  N  I
TATATCTTGGTTTACGATTTATTATAAAAAAAGCATCAAATTCTCGGAATATAAATAATGATAATATAGTCAGAGAAGGAGATTATATATATCTTAATAT 3800

  D  N  I  S  D  E  S  Y  R  V  Y  V  L  V  N  S  K  E  I  Q  T  Q  L  F  L  A  P  I  N  D  D  P  T
TGATAATATTTCTGATGAATCTTACAGAGTATATGTTTTGGTGAATTCTAAAGAAATTCAAACTCAATTATTTTTAGCACCCATAAATGATGATCCTACG 3900
                                        EcoRI
  F  Y  D  V  L  Q  I  K  K  Y  Y  E  K  T  T  Y  N  C  Q  I  L  C  E  K  D  T  K  T  F  G  L  F  G  I
TTCTATGATGTACTACAAATAAAAAAATATTATGAAAAAACAACATATAATTGTCAGATACTTTGCGAAAAAGATACTAAAACATTTGGGCTGTTTGGAA 4000

  G  K  F  V  K  D  Y  G  Y  V  W  D  T  Y  D  N  Y  F  C  I  S  Q  W  Y  L  R  R  I  S  E  N  I  N
TTGGTAAATTTGTTAAAGATTATGGATATGTTTGGGATACCTATGATAATTATTTTTGCATAAGTCAGTGGTATCTCAGAAGAATATCTGAAAATATAAA 4100

  K  L  R  L  G  C  N  W  Q  F  I  P  V  D  E  G  W  T  E  *
TAAATTAAGGTTGGGATGTAATTGGCAATTCATTCCCGTGGATGAAGGATGGACAGAATAATATAATTAAATATTTATTAAAGCTACTTTGATAGGAAAA 4200

ATCAAATTTTATAAAACTTTAAAATAAAAGGTGTGGTTAAATTTTATCTAAATAACTCACTTTATT  4266
```

Fig 10. *Nucleotide sequence of the BoNT/G gene.* The illustrated sequence was derived by amalgamating the nucleotide sequences of the inserts of plasmids pCBG1 to pCBG9 (Fig. 9). The BoNT/G amino acid sequence is given in the single code above the first nucleotide of the corresponding codon.

nucleotide sequences, explaining why probes based on the *botB* sequence have a tendency to cross-react with type G DNA.

1.5 AMINO ACID HOMOLOGIES BETWEEN NEUROTOXINS

Pairwise comparisons of the amino acid sequences of the respective L and H chain components of all currently characterised botulinum neurotoxins and tetanus toxin was undertaken and the results summarised in Table 3. This table does not include comparisons with the BoNT/E isolated from *C. butyricum*, as they are not sufficiently dissimilar from the BoNT/E of *C. botulinum* to warrant individual treatment. The three types of BoNT/F have been included. From this it can be seen that, with notable exceptions, the overall level of identity between the L chains of different toxin serotypes varies from around 30 to 35%. The

HEAVY / LIGHT	A	B	C	D	E	F Proteolytic	F Non-proteolytic	F *C. baratii*	G	TET
A		48	34	35	46	44	42	45	41	35
B	31		39	40	44	38	39	41	55	36
C	32	32		56	36	34	34	33	32	32
D	35	35	47		37	33	37	33	34	36
E	33	33	32	33		68	64	68	41	35
F Proteolytic	32	38	33	35	57		84	79	37	32.5
F Non-proteolytic	32	37	33	35	56	94		73	38	33
F *C. baratii*	32	40	33.5	34	54	63	63		37.5	35
G	33	61	34	37	35.5	38.5	38	40		36
TET	30	50	34	34.5	40	43.5	48	44	44	

Table 3. ***Amino acid homology between the L and H chain components of the different types of BoNT and TeTx.*** Figures represent the % identity between di-chain components. The upper quadrant contains H chain comparisons, the lower L chain homologies. A, B, C, D, E, F, G refer to the respective BoNT, TET represents TeTx. The strains from which the three BoNT/F's were derived were: Langeland, proteolytic *C. botulinum*; ATCC 23387, non-proteolytic *C. botulinum*, and; *C. baratii* ATCC 43756. In the full alignment of Fig. 12 these are labelled, respectively, $BOTF_1$, $BOTF_2$ and $BOTF_3$.

notable exceptions are the degree of sequence identity seen between BoNT/G and BoNT/B (61%), BoNT/F (ATCC 23387) and TeTx (48%), BoNT/E and BoNT/F (57%), BoNT/C and BoNT/D (47%) and BoNT/B and TeTx (50%). The fact that certain BoNT's (BoNT/B,

BoNT/E and BoNT/F) exhibit a greater degree of homology to the TeTx L chain than to other BoNT L chains is particularly striking. With the exception of TeTx, the H chains exhibit a much broader spread of % similarity values than the L chains. The highest degree of similarity is that found between BoNT/E and BoNT/F (68%), closely followed by the 56% similarity found between the H chains of BoNT/C and BoNT/D, and the 55% identity shared by BoNT/B and BoNT/G. The overall disimilarity of the TeTx H chain to BoNT's is consistent with the view that this region is responsible for the essential difference between these neurotoxins, viz, their site of action.

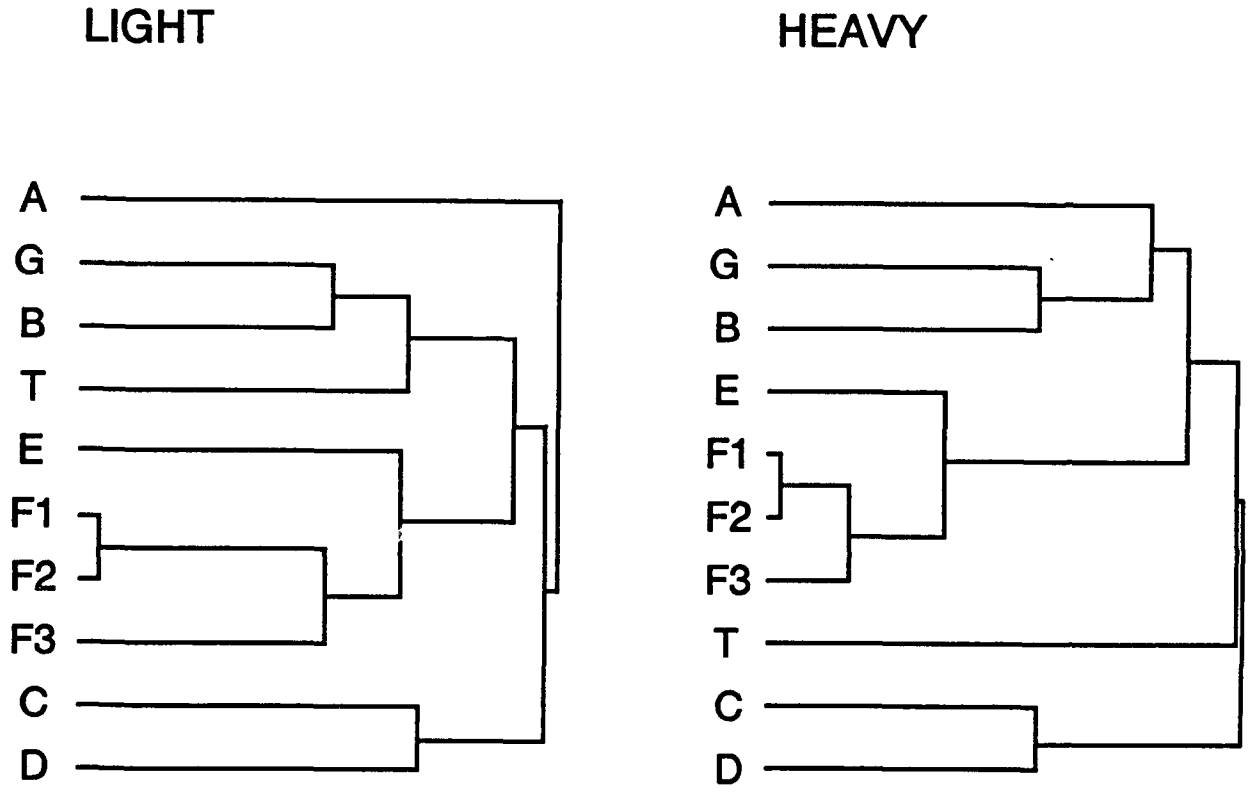


Fig. 11. *Phylogenic relationships between the H and L chains of clostridial neurotoxins.* The distance of the line along the *x* axis is indicative of degree of divergence.

On the basis of L chain comparisons, BoNT/A is the most divergent neurotoxin, exhibiting a low level of homology with all other toxins. The other neurotoxins appear to fall into three groupings, viz, BoNT/B, BoNT/G and TeTx, BoNT/E and BoNT/F, and BoNT/C and BoNT/D. The latter two groups also appear to hold for the H chains, however, in this case BoNT/A falls into the BoNT/G and BoNT/B group, and it is TeTx which shows appears to have no homologous counterpart. These relationships are best illustrated by the phylogenic tree illustrated in Fig. 11. The variance seen in the relative order of relatedness between toxins

dependent of which component of the dichain that is compared is intriguing. It suggest that either L and H chain domains of an individual neurotoxin have evolved at disproportionate rates, or that at various stages during evolution hybrid toxins have arisen by fusion of distinct H and L chain encoding regions.

An alignment of all known neurotoxin sequences, including the three different BoNT/F sequences, but excluding the *C. butyricum* BoNT/E sequence, is presented in Fig. 12. Regions of sequence similarity have been boxed. This demonstrates that the neurotoxins are composed of highly conserved amino acid domains interspersed with amino acid tracts exhibiting little overall similarity. Within the L chain region (average size 440), 63 amino acids are totally conserved. 11 of these conserved amino acids reside in a region (position 216 to 234) which encompasses a histidine rich motif now known to play a role in the zinc endopeptidase cleavage of at least two protein components (dependent on serotype) of the of the putative fusion complex mediating synaptic vesicle exocytosis (Schiavo et al., 1992; 1993; Blasi et al., 1993)

Within the H chain regions (average size 842 amino acids) 93 amino acids are absolutely conserved. Most notable is the high degree of conservation of Trp amino acids. Thus, for instance, of the 11 Trp residues which occur in the BoNT/E H chain, 8 are absolutely conserved in all toxins, while the remaining 3 are conserved in all but one of the neurotoxins at each position. The only Trp that occurs in the BoNT/E L chain is conserved in all neurotoxins. The functional significance of the apparent evolutionary pressure for maintaining this amino acid, or chemically similar residues, at these positions in BoNT and TeTx remains unknown. However, previous studies in which BoNT Trp residues have been selectively modified by chemical means has established a crucial role in both toxicity and immunogenicity (see Dasgupta, 1990). Indeed, in one study the inactivation of a single Trp resulted in near complete detoxification (Shibaeva et al., 1981, cited in DasGupta, 1990). The selective disruption of conserved Trp amino acids in BoNT by site-directed mutagenesis should help identify which residue(s) are important in toxicity and antigenicity.

The most notable tract of sequence divergence between the toxins resides, with the exception of the extreme 10 or so amino acids, in the COOH-termini of the toxins (position 1117 onwards of BoNT/A). Divergence in this latter area would appear consistent with the notion that this domain is involved in BoNT binding, and that the different toxins target different acceptors on the cell surface. The presence of the conserved motif WXFI/VXXXXGW at the extreme COOH-terminus of all neurotoxins (except BoNT/C, where the terminal GW is missing) is especially noteworthy, considering the degree of diversity of the preceding 100 amino acids.

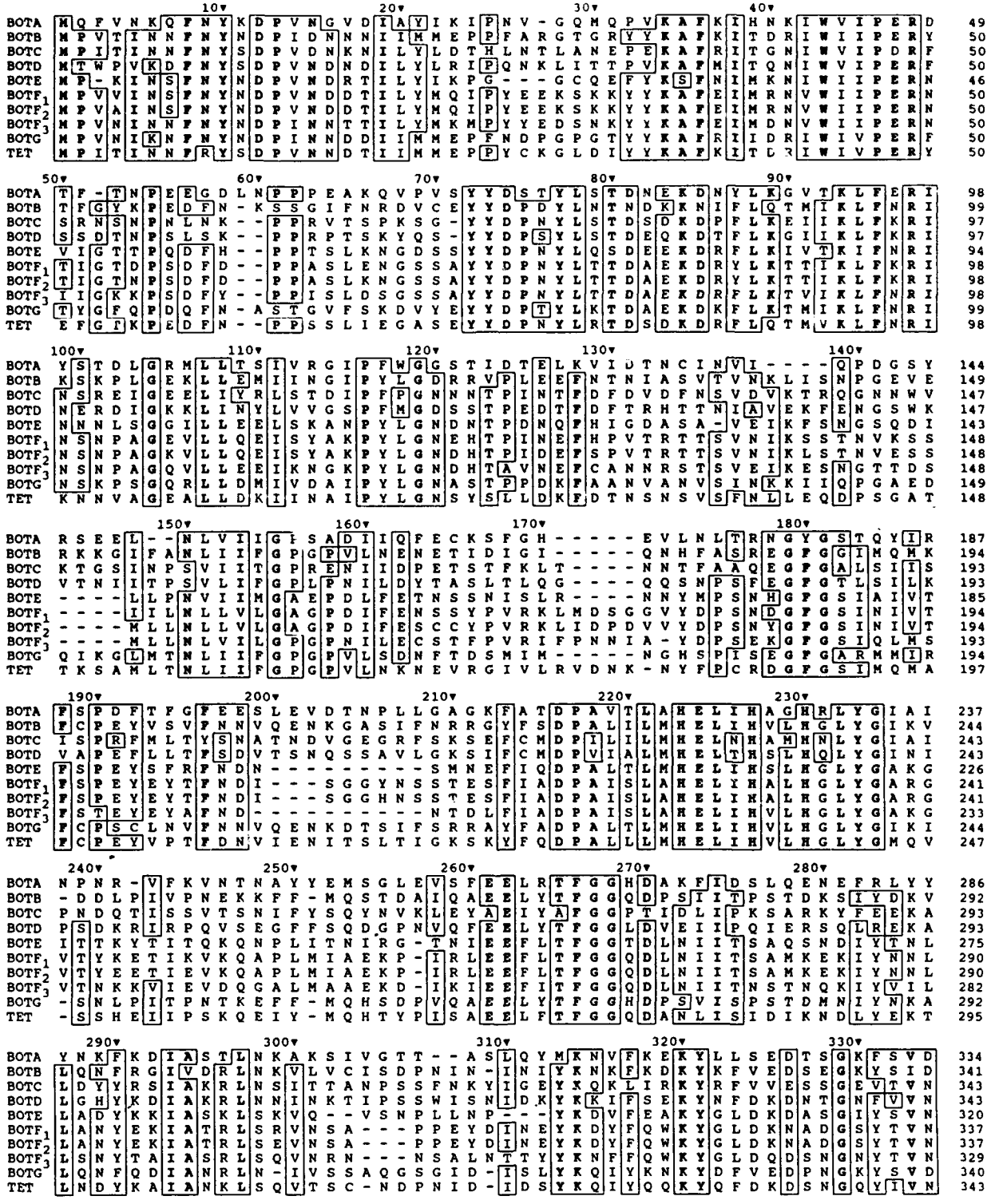


Fig. 12. *Full alignment of all known clostridial neurotoxin sequences.*

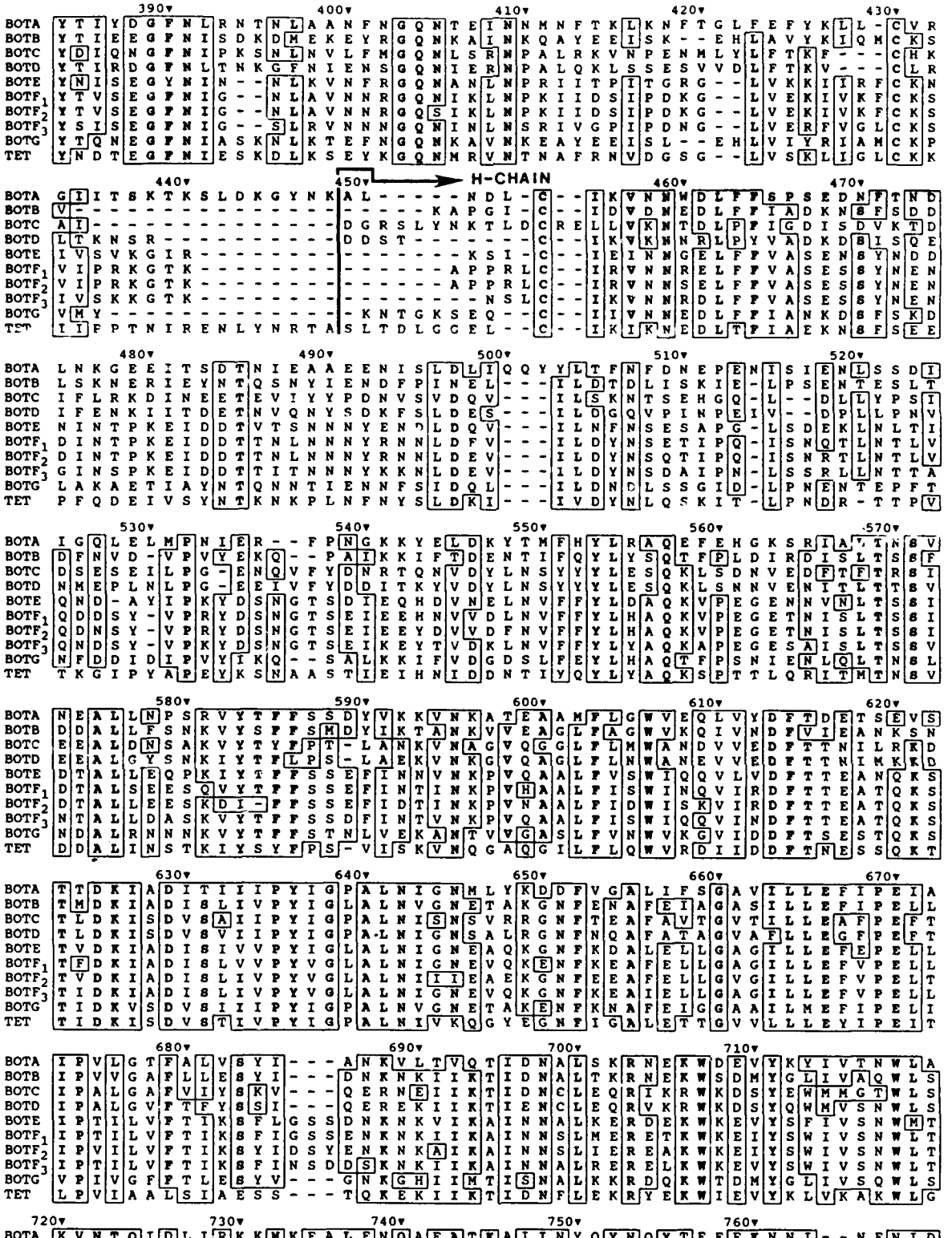


Fig. 12. *Full alignment of all known clostridial neurotoxin sequences.*

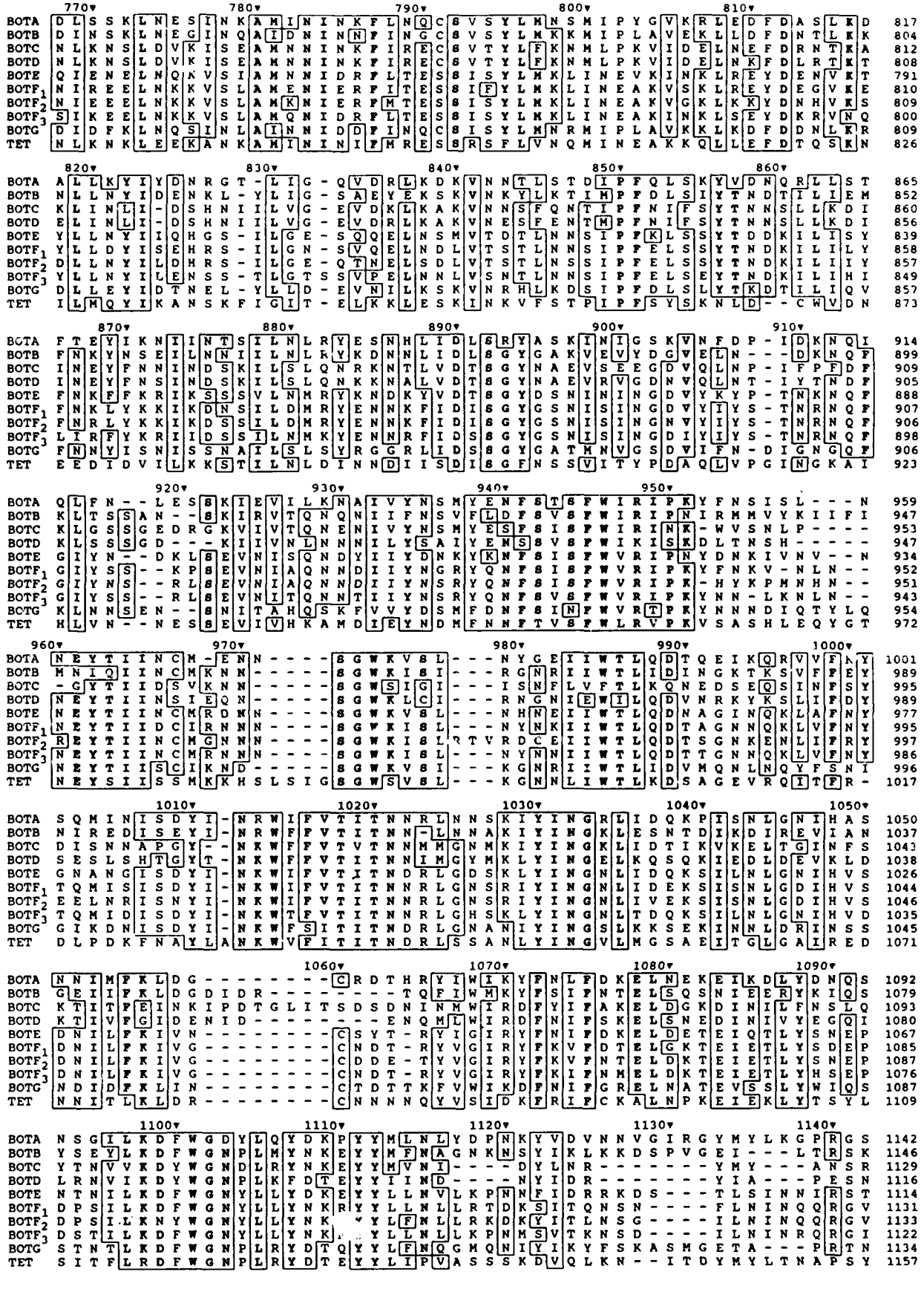


Fig. 12. *Full alignment of all known clostridial neurotoxin sequences.*

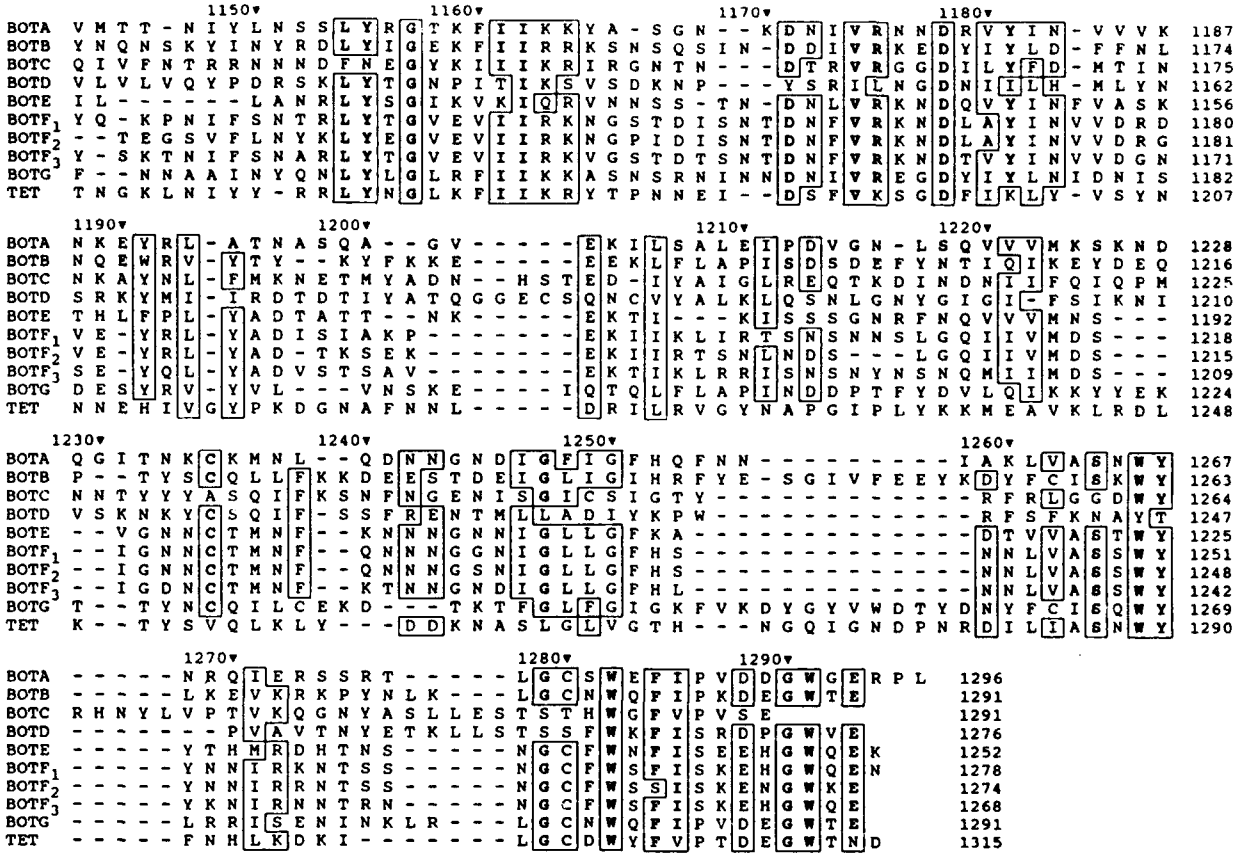


Fig. 12. *Full alignment of all known clostridial neurotoxin sequences.*. The illustrated alignment was essentially derived using the computer programme CLUSTAL (Higgins and Sharp, 1988), and has been gapped to maximise homology. Highly conserved regions have been boxed, and include areas in which conservative replacements have occurred, in addition to sequence identity. Amino acids conserved in at least 8 out of 10 toxins have been emboldened. Numbering above the alignment corresponds to BoNT/A. The Cys amino acids presumed to be involved in the formation of the disulphide bridge between neurotoxin L and H chains are marked by upward facing arrows. $BOTF_1$ = strain Langeland, $BOTF_2$ = ATCC 23387, $BOTF_3$ = ATCC 43756.

The algorithms of Chou and Fassman (1978) and Garnier et al. (1978) were employed to derive predictive representations of BoNT and TeTx secondary structure (data not shown). The results obtained went some way towards confirming the observations of a comparative structural analysis undertaken with purified BoNT/A and BoNT/E (Singh et al., 1990). Thus, the BoNT/E is predicted to contain a lower α-helix content than BoNT/A (BoNT/E, 20%; BoNT/A 27%), and a correspondingly higher content of β-sheet (BoNT/E, 52%; BoNT/A, 46%). No common pattern between the predicted structures of each neurotoxin was, however, apparent. In contrast, a hydrophilicity analysis by the method of Kyte and Doolittle

(1982) demonstrated a high degree of conservation between all 7 neurotoxins in their arrangement of polar and nonpolar amino acids (see Fig. 13). A similar previous analysis of

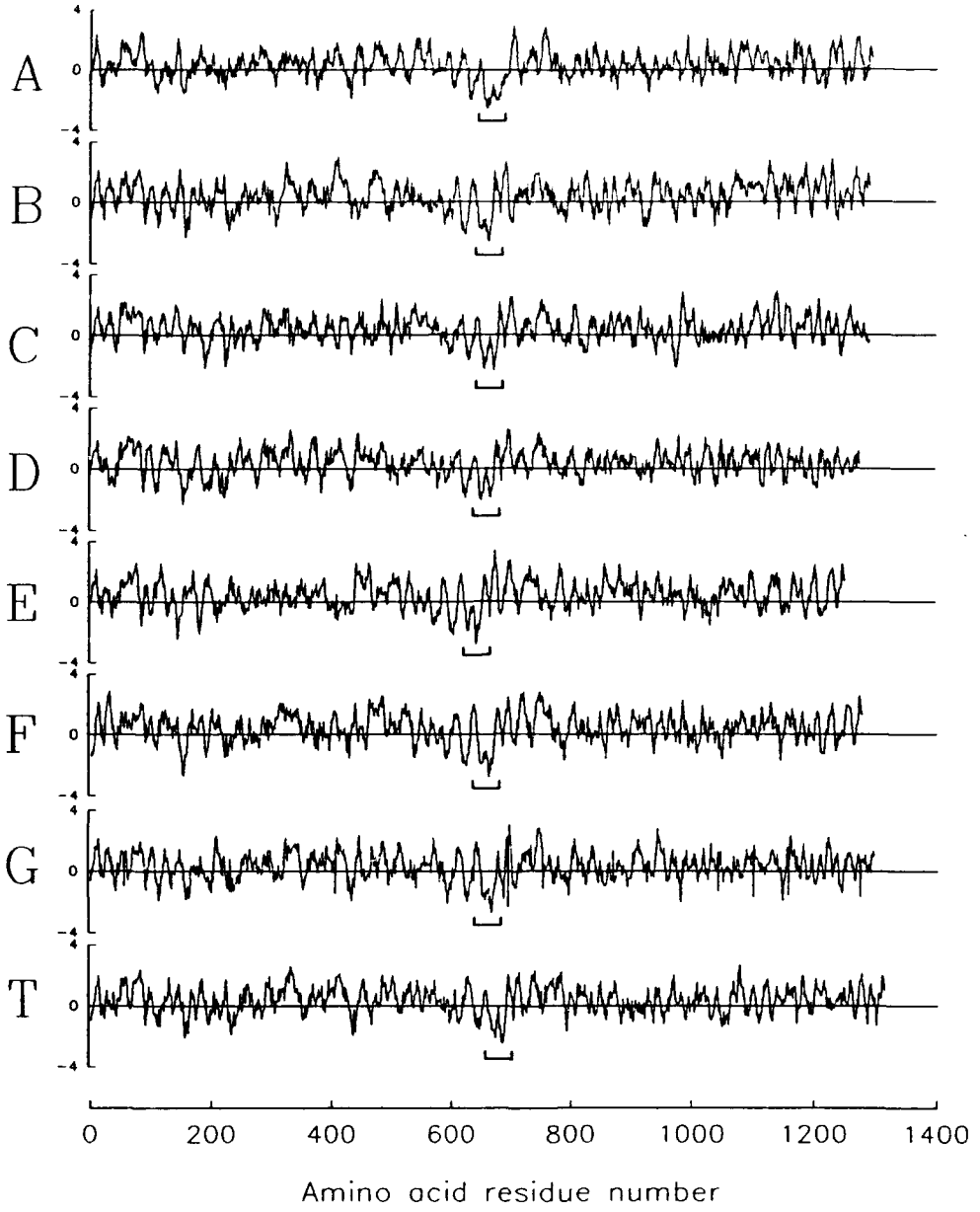


Fig. 13. ***Hydrophobicity plots of all currently characterised clostridial neurotoxins.*** Hydrophobicity was calculated using the computer programme of Kyte and Doolittle (1982) with a window size of 9 amino acids. The average value for each toxin was:- BoNT/A, -0.37; BoNT/B, -0.42; BoNT/C, -0.41; BoNT/D, -0.36; BoNT/E, -0.45, and; TeTx., -0.37. The conserved hydrophobic region is indicated below each profile by a barred line. The respective residues involved are 652 through 687 (BoNT/A), 642 through 671 (BoNT/B), 648 through 678 (BoNT/C), 646 through 674 (BoNT/D), 624 through 654 (BoNT/E), 643 through 673 (BoNT/F), 640 through 669 (BoNT/G) and 660 through 691 (TeTx).

TeTx (Eisel et al., 1986) and BoNT/A (Thompson et al., 1990) had concluded that their H

chains contained a common domain (TeTx, 660 through 691; BoNT/A 652 through 687; Fig. 13) with membrane spanning potential. Use of a synthetic peptide corresponding to this region recently confirmed this conclusion (Wright et al., 1992). The equivalent hydrophobic domains (Fig. 12) are also conserved in BoNT/B (642 through 671) BoNT/E (624 through 654), BoNT/F (643 through 673), BoNT/C (648 through 678), BoNT/D (646 through 674) and BoNT/G (640 through 669).

2. EXPRESSION SYSTEM DEVELOPMENT

2.1 MATERIALS AND METHODS

Bacterial strains, plasmids and culture conditions.

The *E. coli* and *Bacillus subtilis* strains routinely used as the host for recombinant experiments were TG1 (Δ[*lac-pro*] *supE thi hsdD5*/ F'- *traD36 proA*$^+$ *B*$^+$ *lacI*^{q} *lacZΔM15*) and 168 *trpC*, respectively. The *Clostridium acetobutylicum* strain employed was NCIB 8052. The strains of *Clostridium sporogenes* tested were; BM1091, BM1706, BM1758, BM1759, BM1761, BM1763, BM1764, BM1765, BM1767, BM1768, BM1769, BM1774, BM1775, BM1776, BM1780, BM1781, BM1783, BM1784, BM2130 and BM2131. All strains were obtained from Dr. M Hudson, Pathology Division, PHLS CAMR. Recombinant plasmids employed were the pMTL20 series of cloning vectors (Chambers et al., 1988), the replicon cloning vectors pMTL20/21E and pMTL20/21C (Swinfield et al., 1990), pAMB1-derived shuttle vectors pMTL500E/C and pCTC1 (Oultram et al., 1988a; Swinfield et al., 1990; Williams et al., 1990a & b), and the *Clostridium* shuttle vectors pCB3 and pCTC501 (Young et al., 1989a & b).

All clostridial cultures were routinely grown in 2x YTG medium (1.6% tryptone, 1.0% yeast extract, 0.5% NaCl, and 0.5% glucose). In certain instances commercially obtained (Oxoid) reinforced clostridial medium (RCM) was employed, and on other occasions the basal medium of O'Brien and Morris (1971). All manipulations were undertaken under anaerobic conditions using an Anaerobic Work Station Mark III (Don Whitley Scientific, UK). The incubation temperature was routinely 37°C.

Plasmid isolation methodology.

Plasmid DNA was isolated from *E. coli* was as described in 1.2 of this report. Plasmid DNA from clostridial strains was isolated by an alkaline lysis procedure. Cells from a 10ml volume of culture, grown overnight in 2x YTG, were harvested by centrifugation and resuspended in 100μl of 50mM Tris-HCl, 25% (w/v) sucrose, 5 mM EDTA, pH 7.0, and lysozyme added to 10mg/ml. Following an incubation period of 60 min, at 37°C, a 200μl aliquot of freshly prepared 0.2 N NaOH, 1% SDS, was added and the tube inverted before being placed on ice for 5 min. A 150 μl aliquot of ice-cold potassium acetate solution (5M potassium acetate: glacial acetic acid: dH_2O, 60:11.5:28.5)), was added, mixed by vortexing and the tube stored on ice for 5 min. Following centrifugation, for 10 min in a microfuge, the

supernatant was transferred to a fresh 1.5 ml eppendorf tube and an equal volume of phenol/chloroform (1:1) added. After vortexing and centrifugation in a microfuge the upper aqueous layer was carefully removed, mixed with 2 volumes of ethanol and allowed to stand at room temperature for two min. The DNA was precipitated by centrifugation, dried and resuspended in an appropriate volume of TE buffer.

Electroporation

A loopful of fresh culture was used to inoculate 500 μl of 2 X YTG and this was then used to set up dilutions from 10^{-1} to 10^{-6} in 5 ml volumes by serial dilution. Cultures were grown overnight. The two lowest dilutions which had grown were used as inoculum for 100ml 2x YTG which was grown to an OD at 600nm of 0.5 - 0.6 (mid-exponential growth), cooled on ice for a few minutes, then harvested by centrifugation at 5000rpm for 10 min. The cell pellet was washed in 5ml ice-cold electroporation buffer (270 mM sucrose, 1 mM $MgCl_2$, 7 mM sodium phosphate buffer, pH 7.4) and harvested by centrifugation as above. The pellet was finally resuspended in 5ml ice-cold electroporation buffer and held on ice. One μg DNA was added to each cuvette (0.2 cm inter-electrode diameter) followed by 300 μl cell culture. The cuvettes were sealed with plastic insert. A single pulse was delivered: 1.25kV, 100ohms, 25μFD. (Time constant approx. 1.7ms). The culture was removed from cuvette by washing with 1ml 2x YTG and added to a final volume of 3ml of 2x YTG (*ie* a 1 in 10 dilution). A three hour expression period was followed by harvesting by centrifugation as above. The pellet was resuspended in 200 μl of 2x YTG and 100 μl volumes plated on selective agar, containing freshly prepared catalase (final concentration of 400 units/ml).

As far as possible, all manipulations were carried out in an anaerobic cabinet and all media and buffers were allowed to equilibrate in anaerobic conditions overnight. The Biorad "Gene Pulser" was used routinely as the electroporation apparatus.

Conjugation

E. coli cultures were grown overnight to OD at 600nm of >4.0 and *C. acetobutylicum* cultures (mid-exponential phase) to an OD at 600nm of 0.6. The donor and recipient cultures were mixed in a 1000:1 ratio within a total volume of 2ml, passed through a sterile 0.45μm pore size filter (2.5cm in diameter) and the filter was incubated upright overnight on reinforced clostridial medium supplemented with 2 mg of catalase. Growth on the filter was harvested by vortexing in 500 μl 25 mM potassium phosphate pH 7.0, 1 mM $MgSO_4$ and 100 μl volumes were plated on clostridial basal medium supplemented with 10 μg trimethoprim (to select

against *E. coli*) and selective antibiotic. As far as possible all manipulation should be carried out under anaerobic conditions.

Western blotting

E. coli cultures were routinely grown to mid-exponential phase and then induced with IPTG. 1.5ml of bacterial cultures were harvested by centrifugation and resuspended in 300ul PAGE lysis buffer (0.08M Tris-HCL, pH 6.8, 0.1m dithiothreitol, 2% (w/v) SDS, 10% (v/v) glycerol, 0.1mg/ml bromophenol blue) for an OD_{600nm} of 1.5 and boiled for 5 mins. Samples were analysed by electrophoresis immediately or stored at -20°C. SDS polyacrylamide gel electrophoresis was carried out in 10% separating gels with 5% stacking gels run at 100 volts for 4-5 hrs. Pre-stained protein molecular weight markers (Biorad) were used.

After electrophoresis, gels were blotted overnight in Biorad transblot apparatus at 75 volts using Hybond C Extra (Amersham) as the membrane. Use of pre-stained protein markers allowed visualisation of transfer. Following blotting, membrane was incubated with blocking buffer (3% casein in 1xPBS, 0.5% Tween 20) for 45 min. All incubation and washing steps were carried out shaking at room temperature. Membrane was washed twice in PBS Tween then incubated with first antibody diluted in PBS Tween for 90 min. First antibody was either guinea pig *C. botulinum* type A heavy chain anti-sera (from Biologics Division, PHLS CAMR) diluted 1:2000 or goat GST anti-sera (from Pharmacia) diluted 1:2000. The membrane was then washed three times in PBS Tween and incubated with second antibody (anti-guinea pig IgG peroxidase conjugate anti-sera or anti-goat IgG peroxidase conjugate anti-sera, both obtained from Sigma and both diluted 1:2000 in PBS Tween) for 90 min. Nitrocellulose membrane was finally washed four times in PBS Tween prior to peroxidase detection using ECL Western Blotting Kit (from Amersham) according to manufacturer's instructions in association with ECL Hyperfilm (Amersham).

DNA manipulations

Other routine methods of DNA manipulation have been described in section 1.1.

2.2 GENE TRANSFER IN *CLOSTRIDIUM SPOROGENES*

2.2.1 Summary

A total of 20 different strains of *C. sporogenes* have been tested as potential recipients for DNA transfer. Having established that the BioRad Gene Pulser gives the highest rate of electrotransformation in *C. acetobutylicum* (compared to equivalent equipment supplied by Jouan, BTX and Flowgen), attempts were made to transform all strains with a variety of plasmids and differing electrical parameters. Pulses were undertaken at a constant voltage (1.25 kV) and capacitance (25 μFD) but at variable resistance (100, 200 & 400 ohms). Under these conditions the % survival varied from 46 to 13%. Plasmid replicons employed were either from the Gram-positive, broad-host-range plasmid pAMß1, or the *C. butyricum* plasmid pCB101. Selective markers were the *erm* (Em^R) gene of pAMß1 or a *C. perfringens tetP* gene. No transformants were obtained. Attempts to conjugatively mobilise derivatives of these vectors, endowed with the RK2 origin of transfer (*oriT*), from *E. coli* to each strain were similarly unsuccessful.

2.2.2 Results and Discussion

Antibiotic resistance profiles of strains

The successful introduction of an extrachromosomal DNA into bacteria requires that the transformed cell acquires a detectable phenotypic trait. The selectable genetic markers most commonly used are genes specifying resistance to antibiotics. Before attempting to obtain transfer of plasmids into any particular strain of *C. sporogenes*, it was therefore important to establish the antibiotic resistance profiles of the strains to be employed. A 3 ml volume of molten H-top agar was inoculated with 0.1 ml of exponential phase cells (growing in 2 X YTG media) and overlayed onto a 2 X YTG agar plate. When the inoculated agar had solidified, antibiotic-impregnated filter discs (Mast Laboratories Ltd) were placed on the agar surface and the plates incubated overnight at 37°C. The qualitative estimates of zones of inhibition around the different type of disc are indicated in Table 4. These show that, with the notable exception of streptomycin (Sm) and novobiocin (Nc), the 20 strains tested exhibited varying degrees of sensitivity to all the antibiotics tested. Of particular importance was the demonstrable susceptibility of every strain to erythromycin (Em), chloramphenicol (Cm) and tetracycline (Tc). Genes specifying resistance to these three antibiotics form the basis of all currently available clostridial vectors (Young et al., 1989; Rood and Cole, 1991).

Plasmid screening

In parallel to the above tests each strain was screened for the presence of indigenous extrachromosomal elements using a plasmid isolation procedure routinely used in this laboratory for analysing transformants of *C. acetobutylicum* (MATERIALS AND METHODS). The cell lysates obtained were electrophoresed on 1.4% (w/v) agarose gels in

STRAIN	Cm_{25}	Em_5	Fu_{10}	Me_{10}	Nc_5	Pc_1	Sm_{10}	Tc_{25}	Cf_{30}	Mn_5	Cl_2
BM1763	++	++	+++	++	+	++	-	++	+++	++	++
BM1758	+++	+++	++	++	+	++	-	+++	+++	++	++
BM1767	++	++	+++	++	-	++	-	++	+++	++	++
BM1774	++	++	+++	++	+	++	-	+++	+++	++	++
BM1796	++	++	+++	++	+	++	-	+++	+++	++	++
BM1781	++	++	+++	++	+	++	+	+++	+++	++	++
BM1775	+++	++	+++	++	+	++	-	+++	+++	++	+
BM2131	+++	++	+++	++	+	++	-	+++	+++	++	+
BM1780	++	++	+++	++	+	+++	+	+++	+++	++	++
BM1759	+	+++	+++	+++	-	-	-	++	+++	+++	++
BM1706	-	+++	+++	+++	-	++	-	++	+++	++	++
BM1764	++	++	+++	++	+	+++	-	+++	+++	++	++
BM1776	++	+++	+++	++	-	++	-	+++	+++	++	+
BM1783	+++	++	+++	++	-	++	-	++	++	+++	++
BM1768	+++	++	+++	++	++	++	-	+++	+++	++	-
BM1761	++	++	+++	++	-	++	-	+++	+++	++	+
BM1091	++	++	+++	++	-	++	-	+++	+++	++	+
BM2130	+++	++	+++	++	-	+++	-	+++	+++	++	+

Table 4. *Susceptibility of C. sporogenes strains to various antibiotics*

- = no zone of inhibition; + = zone up to 10 mm in diameter; ++ = zone of 11-20 mm; +++ zone >20 mm (all zones include the disc diameter of 6.5 mm). Antibiotic abbreviations are Cm, chloramphenicol; Em, erythromycin; Fu, fusidic acid; Me, methicillin; Nc, novobiocin; Pc, penicillin; Sm, streptomycin; Tc, tetracycline; Cf, cefoxitin; Mn, metronidazole, and; Cl, clindamycin. Antibiotic concentrations are given in subscripts , following each abbreviation, in μg per ml.

addition to the standard 0.8% (w/v) gels normally employed in plasmid analysis. This higher concentration of agarose ensures that any circular DNA species, "masked" by chromosomal DNA on a 0.8% gel, migrates substantially slower than chromosome and is therefore easily

visualised (Minton et al., 1983a). No evidence for the presence of plasmid DNA was found in the lysates of any of the 20 strains. In a further series of experiments the methods of Roberts *et al.* (1986) and Weickert *et al.* (1986), were employed. These procedures have previously been used to detect plasmid DNA in *Clostridium perfringens* and *Clostridium absonum*, and in *Clostridium botulinum* Type A strains, respectively. Although both methods proved applicable to a control *C. acetobutylicum* NCIB 8052 culture containing pCB3, no plasmids were detected in the lysates of any of the *C. sporogenes* strains under investigation.

Evaluation of various electroporators

Since the development of our original procedure for effecting the introduction of plasmid DNA into *C. acetobutylicum* using a BioRad Gene Pulser (Oultram et al., 1988a), a number of other manufacturers have brought alternative machines onto the market. It was therefore considered timely to undertake a comparative evaluation of more recent apparatus, on the assumption that an increase in transformation frequencies may accrue. Three such machines were tested, alongside the BioRad Gene Pulser, for their efficiency in transforming *C. acetobutylicum* NCIB 8052 with plasmid pMTL500E (see Fig. 16). The BTX electroporator may be essentially viewed as equivalent in specification to the BioRad apparatus. The Jouan electropulser differs from other commercially available apparatii in that it generates a square wave pulse, which theoretically provides a constant field during discharge. The Flowgen Cellject resembles the BioRad and BTX machine, in that it discharges an exponential wave, but differs in the facility for discharging a preprogrammed second pulse, immediately after the first.

Electroporator	Transformation Frequency (per μg DNA)
BioRad Gene Pulser	1.2×10^2
BTX Electroporator	0.8×10^2
Flowgen Cellject	0.5×10^2
Jouan Electropulser	0

Table 5. *C.acetobutylicum transformation frequencies employing different electroporation apparatus.*

Each machine was tested over a range of pulse parameters. With those machines that did mediate transformation, however, these parameters were essentially equivalent to those (1.25 kV, 100 ohms, 25 μFD) which gave the highest levels of DNA transfer with the routinely used BioRad Gene Pulser, viz., identical for the BTX, and 1.25 kV, 90 ohms and 40 μFD for the

Celljet. In the case of the Jouan Electropulser no transformants were obtained under any of the conditions employed. Indeed, the machine appeared incapable of effecting DNA transfer even into *E. coli*. This failure would appear to have been largely due to the ineffective electroporation chamber supplied with the apparatus, which was cumbersome to use and suffered from sample leakage. The other two machines both proved effective in eliciting transformation of *C. acetobutylicum* NCIB 8052 (Table 5). However, under optimum conditions, use of the BioRad machine consistently resulted in the highest transformation frequencies. The subjection of cell suspensions to a second pulse, of various magnitudes, using the Cellject gave a slight increase (c. 10%-20%) in the number of transformants, but the frequency obtained was significantly lower than those achieved with the BioRad apparatus. It was concluded that the electroporators of other manufacturers offered no advantage over the BioRad Gene Pulser, and this apparatus was used in all subsequent experiments with *C. sporogenes*.

Attempted electrotransformation of strains of C. sporogenes.

Prior to attempting the transformation of any particular strain of *C. sporogenes*, it was of interest to estimate the effect of electrical pulses on cell viability. Cell suspensions, prepared as for *C. acetobutylicum*, were therefore divided in two, and one fraction subjected to pulses of various magnitudes before serial dilutions of both cell fractions were plated onto 2 X TYG agar. From the viable colony count obtained with the two cell fractions it was possible to estimate the % cell survival after each pulse. Some representative data is shown in Table 6.

STRAIN	% SURVIVAL		
	100 ohms	200 ohms	400 ohms
BM1781	46	19	16
BM2131	40	22	14
BM1706	45	20	21
BM1759	38	18	15
BM1091	37	19	13
NCIB 8052	8.5	3.5	1.05

Table 6. ***Percentage survival of C. sporogenes cells, compared to C. acetobutylicum NCIB 8052***

The results obtained indicated that all the *C. sporogenes* strains under investigation exhibited a similar levels of fragility with respect to the pulse applied. It was further apparent *C. sporogenes* is generally a more robust organism than *C. acetobutylicum*.

These experiments established that the field strength applied was having some effect on cell viability. However, as there are no hard and fast rules as to the level of cell survival most appropriate for successful transformation, attempts to transform the 20 strains of

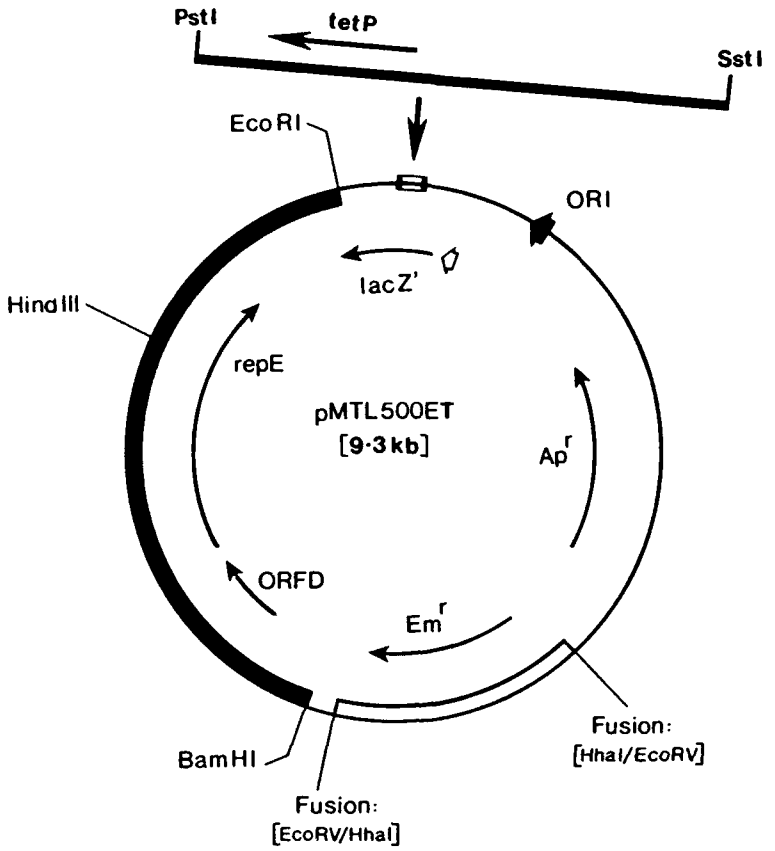


Fig. 14. ***The Clostridium/ E.coli shuttle vector pMTL500ET.*** Constructed by isolating a 2.9 kb *Sst*I-*Pst*I fragment encoding *tetP* from the *C. perfringens* plasmid pJIR71 (Rood and Cole, 1991) and inserting it between the equivalent sites of pMTL500E (Oultram et al., 1988a).

C. sporogenes were undertaken using a pulse of constant voltage (1.25 kV) and capacitance (25 μFD), but at the three different resistances employed in the cell viability experiments, viz., 100, 200 and 400 ohms. The plasmids employed were, pCB3 (Young et al., 1989), pMTL520 (Minton et al., 1990a) and pMTL500ET (Fig. 14). Plasmid pMTL500ET is based on the replicon of the *Enterococcal faecalis* plasmid pAMß1, widely recognised as possessing an extremely broad host range amongst Gram-positive bacteria. Plasmids pMTL520 and pCB3 utilise the replicon of the *Clostridium butyricum* plasmid pCB101 (Minton and Morris, 1981). The selective marker of pCB3 is the pAMß1 *erm* gene (Em^R), that of pMTL520 the *Clostridium perfringens tetP* gene (Tc^R), while pMTL500ET specifies both resistance genes.

In the vast majority of cases, no *C. sporogenes* colonies resistant to either Tc or Em were obtained. A number of putative transformants did result from experiments involving BM1769,

1776, 1783 and 1706, and the plasmid pMTL500ET, and BM1783 and the plasmid pCB3. Subsequent small scale isolation procedures undertaken on representative colonies failed to reveal the presence of extrachromosomal DNA in the resultant cleared lysates. Furthermore, the lysates from the putative pMTL500ET were incapable of transforming competent *E. coli* to Ap^R. In contrast, 8 Ap^R transformants were obtained using the BM1783 lysate derived from the putative pCB3 transformant. Although all 8 *E. coli* transformants were shown to contain plasmid DNA, only 1 gave a restriction pattern characteristic of pCB3. In further tests, radiolabelled pCB3 DNA was used in a Southern blot experiment against total DNA isolated from the putative BM1783 transformant. No positive signal was detected.

Attempts to obtain further transformants of either of the 5 *C. sporogenes* strains proved unsuccessful. This included experiments in which the strains were grown in media containing 2% glycine, prior to the preparation of "competent" cell suspensions. Electrotransformation as a means of eliciting DNA transfer was therefore abandoned in favour of conjugative procedures.

Conjugative DNA transfer

The ability of IncP plasmids to effect the mobilisation of co-resident cloning vectors from an *E. coli* donor to a variety of Gram-positive recipient is now well documented (Trieu-Cuot et al., 1987). Indeed, previous studies have shown that when pMTL500E or pCB3 is endowed with the transfer origin of the IncPII plasmid RK2 (*oriT*), then conjugative transfer of the resultant plasmids (pCTC1 and pCTC501, respectively) was demonstrable between a Tra^+ (RK2) *E. coli* donor and *C. acetobutylicum* NCIB 8052 (Williams et al., 1990a & b). To test the applicability of this method to *C. sporogenes* all 20 strains were used as recipients in filter matings using the Tra^+ *E. coli* host SM17 containing either pCTC1 or pCTC501. Strains were examined in batches of 5, and every experiment included a filter mating employing *C. acetobutylicum* NCIB 8052 as the control. In no instance were any Em^R colonies recovered from a mating involving a *C. sporogenes* as the recipient. In contrast, in every batch of matings, the *C. acetobutylicum* control experiment consistently gave Em^R transconjugants.

2.3 AN EXPRESSION SYSTEM FOR *CLOSTRIDIUM ACETOBUTYLICUM*

2.3.1 Summary

The inability to effect the transfer of plasmid DNA to any strain of *C. sporogenes* led to the adoption of *C. acetobutylicum* NCIB 8052 as the proposed host for production of BoNT toxoid. Efforts focused on deriving a regulated expression system based on the previously constructed *fac* promoter, composed of the transcriptional initiation signals of the *C. pasteurianum* ferredoxin gene in which a synthetic *lac* operator sequence has been inserted immediately 3' to the +1 nucleotide. As it was shown that transcription from *fac* can be regulated by the *lacI* gene in *E. coli*, efforts concentrated on attempts to obtain *lacI* expression in *C. acetobutylicum* NCIB 8052. These experiments revolved around the use of a *lacI* gene derivative which had been transcriptionally coupled to a Gram-positive vegetative promoter, that of the *B. subtilis vegII* gene. Attempts to construct a second plasmid compatible with pMTL500F, into which *lacI* could be inserted, could not be undertaken as no alternative selectable marker to that (*erm*) carried by pMTL500F could be found. Possible candidates examined included Gram-positive genes specifying resistance to Tc, Ap and Cm. A strategy was formulated whereby a replication impaired plasmid (pMTL513E) was employed to bring about the integration of the *lacI* gene into the chromosomal *gutD* gene of a Leu$^-$ mutant of NCIMB 8052. Selection for insertion of *lacI* was made possible by the co-integration of a clostridial *leuB* gene, converting the host to prototrophy. The *fac* promoter of pMTL500F was not, however, regulated in cells carrying the integrated *lacI* gene. Subsequently the *lacI* gene was succesfully introduced into the backbone of pMTL500F to give pMTL500FI. A promoter-less copy of a *cat* gene was introduced into pMTL500FI. Expression of *cat* from the resultant plasmid appeared to be constitutive in both *E. coli* and *C. acetobutylicum*. In *B. subtilis*, however, expression levels were induced between 2- to 5-fold, with fully induced cells producing CAT at up to 20% of the cell's soluble protein.

2.3.2 Results and Discussion

Transcription from the clostridial fac promoter is regulated by LacI

The failure to achieve demonstrable DNA transfer into any of the *C. sporogenes* strains tested necessitated the use of *C. acetobutylicum* as an alternative host. This clostridial species has a number of advantages over *C. sporogenes*. On a practical level, we have already

developed the necessary means of manipulating this species. Equally as important, this species has no known association with human disease and should therefore command a lower Access factor in any proposed recombinant experiments. The proposed expression of BoNT gene subfragments can therefore be undertaken at a lower category of containment. Furthermore, parallel studies undertaken in this laboratory have resulted in the construction of an expression cartridge, similar to that proposed for the *C. sporogenes rrn* promoter, based on the transcriptional signals of the ferredoxin (Fd) gene of *Clostridium pasteurianum* (Minton et al.,

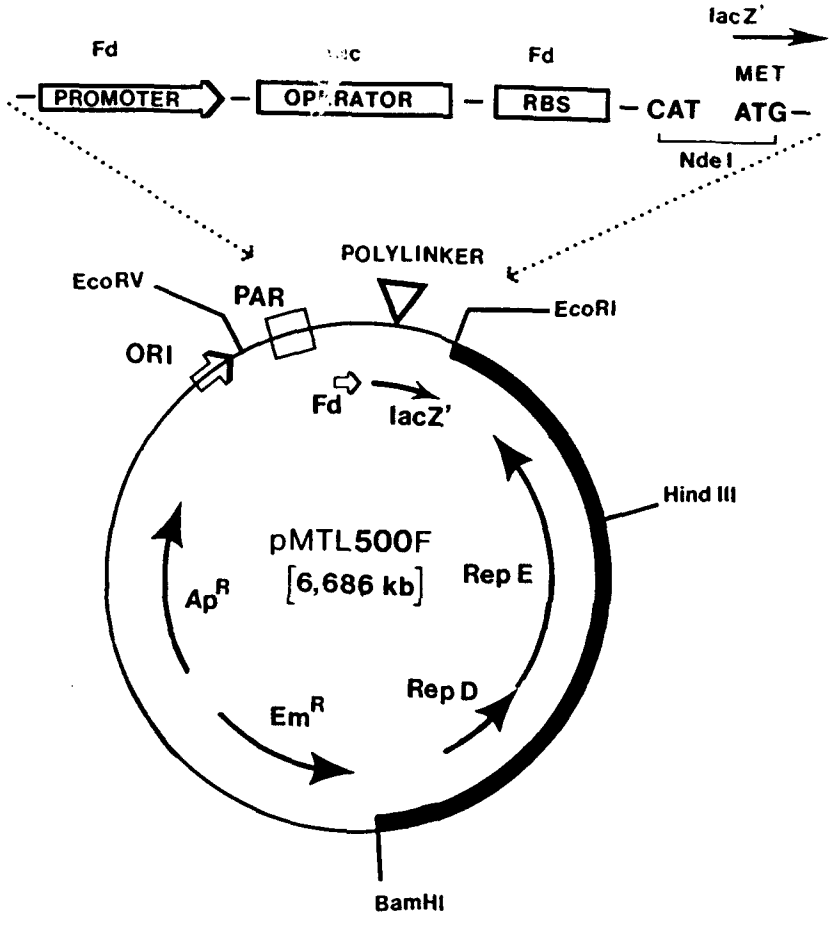


Fig. 15. *The Clostridium acetobutylicum expression vector, pMTL500F.* Plasmid pMTL500F was constructed by replacing the *lac po* region of pMTL500E with the indicated modified (see Minton et al., 1990a) Fd promoter. During its derivation, plasmid pMTL500F also acquired the pSC101 stability function, *par* (PAR). The ATG tri-nucleotide of the indicated *Nde*I restriction recognition site corresponds to the AUG translational start codon of *lacZ'*, and is immediately preceded by the Fd ribosome binding site (RBS). The multiple cloning sites (MCS) are those of pMTL20 (Chambers et al., 1988).

1990a). This expression cartridge was shown to direct the expression of the pC194 *cat* gene such that the encoded protein represented between 3 and 7% of the cells' soluble protein (Minton et al., 1990b) In more recent studies this promoter has been modified by the precise insertion of an *E. coli lac* operator sequence at the Fd +1, and the resultant promoter

derivative (designated *fac*) inserted into pMTL500E in place of the natural promoter of the *lacZ'* gene. Thus, in the derived plasmid, pMTL500F (Fig. 15), expression of *lacZ'* is under the transcriptional control of *fac* (Minton et al., 1990b).

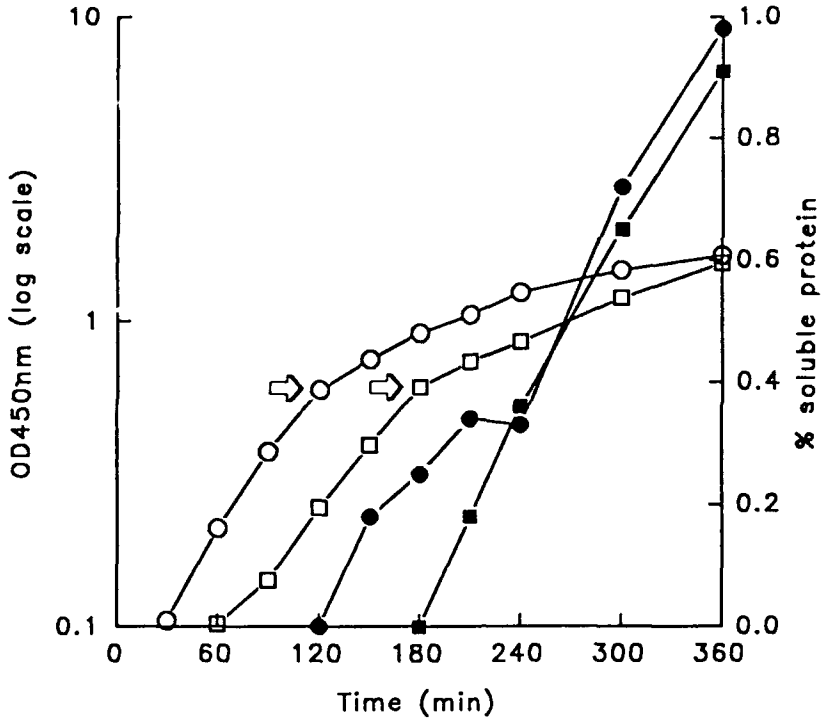


Fig. 16 ***Inducible expression of the pC194 cat gene cloned in pMTL500E and pMTL500F.*** A promoterless copy of the pC194 *cat* gene, excised from pMTL20C (Swinfield et al., 1990) as a 0.8 kb *MnlI* fragment, was inserted into the *SmaI* site of pMTL500E and pMTL500F, such that transcription was dependent on the *lac* or Fd promoter, respectively. The two recombinant plasmids were independently introduced into *E. coli* TG1 containing the $lacI^q$-encoding plasmid pNM52 (Gilbert et al., 1986), and the two clones grown in 2XYT broth to an OD_{450} of 0.6. At this point expression was induced by addition of IPTG (indicated by an arrow) to a final concentration of 1 mM. CAT activity of cells carrying pMTL500E (●) or pMTL500F (■) is expressed as % cell soluble protein. The culture OD_{450} of cells carrying pMTL500E and pMTL500F is indicated by (○) and (□), respectively.

The presence of the *lac* operator should enable transcription from *fac* to be blocked by binding of the LacI protein. Derepression may subsequently be achieved by the addition of the inducer IPTG. Such inducibility requires that the *lacI* gene is efficiently expressed in the recombinant host employed. In our preliminary studies the pC194 *cat* gene was inserted into pMTL500F and the resultant plasmid introduced into an *E.coli* host which carried the *lacI* gene on a co-resident, compatible plasmid, pNM52 (Gilbert et al., 1986). When cells carrying both plasmids were grown overnight in the presence or absence of IPTG, significant repression of *cat* expression was evident. Thus, non-induced cells synthesised CAT to levels of approx.

1.0% of the cells' soluble protein, compared to the 13% levels attained in cells supplemented with IPTG.

A clearer idea of the degree of repression/inducibility was obtained by undertaking the experiment outlined in Fig. 16. Cells carrying pNM52 and either pMTL500E*cat* or pMTL500F*cat* were grown in rich media to mid exponential phase when IPTG was added to both cultures, at a final concentration of 1.0 mM. It can be seen that prior to induction no CAT activity was detectable. Following the addition of IPTG, rapid induction of *cat* expression was evident. Most encouragingly the degree of repression/ induction exhibited by the natural *lac* promoter (pMTL500*cat*) and the *fac* promoter (pMTL500F*cat*) was identical.

Towards expression of lacI in C. acetobutylicum

Having established that *fac* can be regulated by LacI repressor protein, efforts focused on effecting expression of this gene in *C. acetobutylicum* NCIB 8052. Previous workers have elicited expression of *lacI* in the Gram-positive bacterium *B. subtilis* by coupling transcription to a *Bacillus* vegetative promoter and inserting the modified gene either, into the backbone of the expression vector itself (pREP8), or into a second compatible plasmid (LeGrice et al., 1987). Therefore, initially we attempted to insert a pREP8-derived *lacI* encoding DNA fragment into the specially constructed unique *Eco*RV site of the expression vector pMTL500F. Accordingly a 1.4 kb *Eco*RI-*Pvu*I fragment carrying *lacI* was excised from pREP8), blunt-ended by treatment with T4 DNA polymerase and ligated to *Eco*RV cleaved pMTL500F. Subsequent analysis of the recombinant plasmids obtained, however, indicated that severe structural rearrangements had occurred.

The alternative strategy of inserting this gene into a second co-resident plasmid requires the availability of a plasmid which is not only compatible with regard to replication apparatus (ie., different replicon), but in addition, to prevent intermolecular recombination, should not possess DNA homology. We have previously constructed (Minton et al., 1988) such a prototype vector(pMTL520) which, with reference to pMTL500F, fulfils all these criteria. Thus whereas pMTL500F is based on the *E. coli* ColE1 replicon, pMTL520 utilises the p15a replicon. Similarly, pMTL500F uses the pAMß1 replicon and *erm* gene, whereas pMTL520 makes use of the pCB101 replicon and *tetP* from a *C. perfringens* R-factor. However, repeated attempts to transform *C. acetobutylicum* to Tc^R (10 μg/ml) were unsuccessful, raising doubts as to the suitability of pMTL520 for use in *C. acetobutylicum*.

The tetP gene cannot be used as a selective marker in C. acetobutylicum

Two explanations may be evoked to explain the inability of pMTL520 to transform *C. acetobutylicum*. Either: (i) although pMTL520 confers resistance to Tc on *E.coli* hosts, the *tetP* gene does not function in *C.acetobutylicum*, or; (ii) the pCB101 replicon became inactivated during the construction of the vector. To clarify the situation a second plasmid was constructed by inserting the *tetP* gene into pMTL500E (Fig. 14). This new plasmid, pMTL500ET, therefore encodes both *erm* and *tetP*. Confirmation that both antibiotic resistance genes function in a Gram-positive host was obtained by transforming *B. subtilis*, where it proved possible to select for transformants either on the basis of Em^R of Tc^R. Transformation of *C. acetobutylicum* was then repeated using pMTL500ET DNA with selection on plates either containing Em (10 μg/ml) or Tc (10 μg/ml). Transformants were only obtained on the former plates. Furthermore these Em^R transformants subsequently failed to grow on agar medium containing 10 μg/ml Tc.

TETRACYCLINE CONCENTRATION	GROWTH of NCIB 8052	
	Plasmid-free	+ pMTL500ET
0	+++	+++
1 μg/ml	-	+++
2.5 μg/ml	-	++
5 μg/ml	-	+
10 μg/ml	-	-

Table 7. *Growth of NCIB 8052 and a pMTL500ET transformant on media supplemented with Tc*

The inability of pMTL500ET Em^R transformants to grow on plates containing Tc prompted an examination of the level of susceptibility of *C. acetobutylicum* to this antibiotic over a range of concentrations. The results are illustrated in Table 7. *C. acetobutylicum* was found to be incapable of growth at levels as low as 1 μg/ml. In contrast, a pMTL500ET transformant (selected on the basis of resistance to Em) was capable of normal growth at this level of antibiotic, reduced growth at Tc concentrations of 2.5 μg/ml and sparse growth on agar containing 5 μg/ml Tc. The transformation experiment with pMTL500ET was therefore repeated, with selection on plates containing 1.0 and 2.5 μg/ml of Tc. Although Tc^R colonies were obtained at both concentrations, an almost equivalent number were obtained using cells which received no plasmid DNA. Furthermore, replica plating of the putative transformants

onto agar media supplemented with Em revealed that no colony had become Em^R. Raising the level of Tc in agar plates to 4 μg/ml appeared to prevent any growth of cells which did not receive plasmid DNA. However, the low number of colonies obtained from cells treated with pMTL500ET DNA were all found to be Em^S, indicating they were not true transformants. Indeed, no extrachromosomal DNA was evident when appropriate cleared lysates were analysed by agarose gel electrophoresis. It was concluded that, although *tetP* appears to confer Tc^R on *C. acetobutylicum* once the plasmid carrying it has become established in the cell, it is not possible to directly select for Tc^R in transformation experiments. This was confirmed at a later stage in the project when a pAMß1-based plasmid encoding *tetM* obtained from Peter Durre at Gottingen, FRG. was found to be unable to transform NCIB 8052 to Tc^R.

Alternative selective markers

Because the availability of only one selective marker (*erm*) places severe limitations on any future recombinant manipulations in *C. acetobutylicum*, the elucidation of a second marker is a matter of some importance. Reliance on commonly used genes specifying resistance to Cm and Km have previously proven inappropriate for this *Clostridium* spp. (see Oultram et al., 1987). Some authors have circumvented the problems associated with the anaerobic reduction of chloramphenicol by using thiamphenicol, eg., in *Clostridium thermohydrosulfuricum* (Soutschek-Bauer et al., 1985). The possibility of using this analogue for selection of pMTL500F*cat* transformants was therefore explored. As growth experiments demonstrated that *C. acetobutylicum* NCIB 8052 grew on levels of thiamphenicol up to and including 100 μg/ml, a concentration of 150 μg/ml was used in selective plate. However, although pMTL500F*cat* electrotransformants could be readily selected on the basis of Em^R, no colonies were obtained on the plates containing thiamphenicol.

During the course of this work a vector based on the *C. perfringens* plasmid pIP404 was constructed by Julian Roods laboratory which encodes both *erm* and *cat* (Sloan et al., 1992). Interestingly, when *C. acetobutylicum* was transformed with this plasmid, pJIR418, colonies were obtained at equal frequencies on agar plates containing either Cm or Em. Furthermore, all Em^R colonies were also Cm^R, and *vice versa*. However, no extrachromosomal DNA could be detected in *C. acetobutylicum* lysates and lysate aliquots failed to transform either *E. coli* or *B. subtilis* to Em^R or Cm^R. To circumvent the apparent inability of the pIP404 replicon to function in *C. acetobutylcium*, the pJIR418 *cat* gene was excised as a blunt, 1.3 kb *Sma*I-*Nae*I fragment and converted to a sticky-ended fragment by passage through the pMTL21 polylinker region, and reisolating it as a *Sst*I-*Bam*HI fragment. This fragment was then cloned into the equivalent sites of the *Clostridium* shuttle vector pMTL500E. The vector obtained, pMTL500EC, was then transformed into *Clostridium acetobutylicum* with selection

for either Em^R or Cm^R colonies. All the Em^R colonies obtained were shown to be Cm^R. Only 60% of the Cm^R transformants, however, had also become Em^R. Furthermore, it was noticeable that significantly higher numbers of "transformants" were obtained on Cm plates than on Em plates.

A further potential marker gene that could be employed is a gene specifying resistance to ampicillin. Such a determinant encoding a typical "pBR322-like" ß-lactamase from a *Staphylococcus aureus* plasmid (pS1) has recently been sequenced (East and Dyke, 1989). However, although the sequenced *bla* gene alone is sufficient for Ap^R in *E. coli*, a region of DNA 5' to the gene is required for resistance in staphylococci. A plasmid carrying the whole determinant necessary for Ap^R in a Gram-positive bacterium, pAE306, was obtained from Dr K Dyke at Oxford and a 4.0 kb fragment excised and inserted into pMTL520. Although the resultant plasmid conferred Ap^R on an *E. coli* host, Ap^R transformants of *C. acetobutylicum* were not obtained. Subsequent dialogue with Dr Dyke's laboratory indicated that rearrangements of the Ap^R determinant of plasmid pAE306 had occurred. A second plasmid was therefore obtained, pSLJ104, which carried the entire Tn552 transposon, encompassing *blaZ*, on a 6.0 kb *Bam*HI fragment. However, attempts to insert this fragment into the polylinker site of pMTL500E consistently resulted in recombinant plasmids in which structural rearrangements/ deletions were apparent.

The final selective marker examined was the *C. pasteurianum leuB* gene. A 2.2 *Cla*I-*Sph*I fragment carrying this gene was previously cloned into pMTL500E and the resultant plasmid, pLEU100, transformed into a leucine auxotroph of *C. acetobutylicum*, SA9. All of the Em^R transformants obtained were restored to prototrophy (Oultram et al., 1988). It was therefore of interest to repeat this experiment, but in contrast select directly for Leu^+ colonies, ie., test whether *leuB* can be used as a primary selectable marker, as in *Saccharomyces cerevisiae*. However, no colonies were obtained when SA9 cells were electroporated with pLEU100 DNA and plated on clostridial basal medium containing no leucine. All transformants selected on the basis of Em^R were, however, Leu^+. Thus, it is not possible to directly select for acquisition of *leuB*, but it can be employed in secondary selection once phenotypic expression has occurred.

The gut operon as a potential site for homologous integration

The facility for effecting the insertion of heterologous DNA into the host chromosome, by homologous recombination, offers considerable advantages in any proposed programme of strain manipulation. The principal attraction is that it circumvents the problems of recombinant segregational instability commonly associated with autonomous vectors. Thus,

the ability to integrate genes into the *C. acetobutylicum* chromosome offers great potential in the future generation of strains expressing *bot* gene subfragments. Such technology, however, also provides the facility for generating a strain in which *lacI* has been inserted into the chromosome.

Integrative strategies require two components: (i) a cloned region of the host chromosome, to provide homology for recombination, the disruption of which is not deleterious to cell growth, and; (ii) a vector delivery system, the replication properties of which favour integration. We have just completed sequencing the *gut* operon of *C. acetobutylicum*, which encodes the genes necessary for glucitol (sorbitol) transport/metabolism. The operon (Fig. 17) has the same overall arrangement as *E.coli* (Yamada and Saier, 1987), but additionally contains a gene coding for a protein exhibiting homology to the ORF U polypeptide of

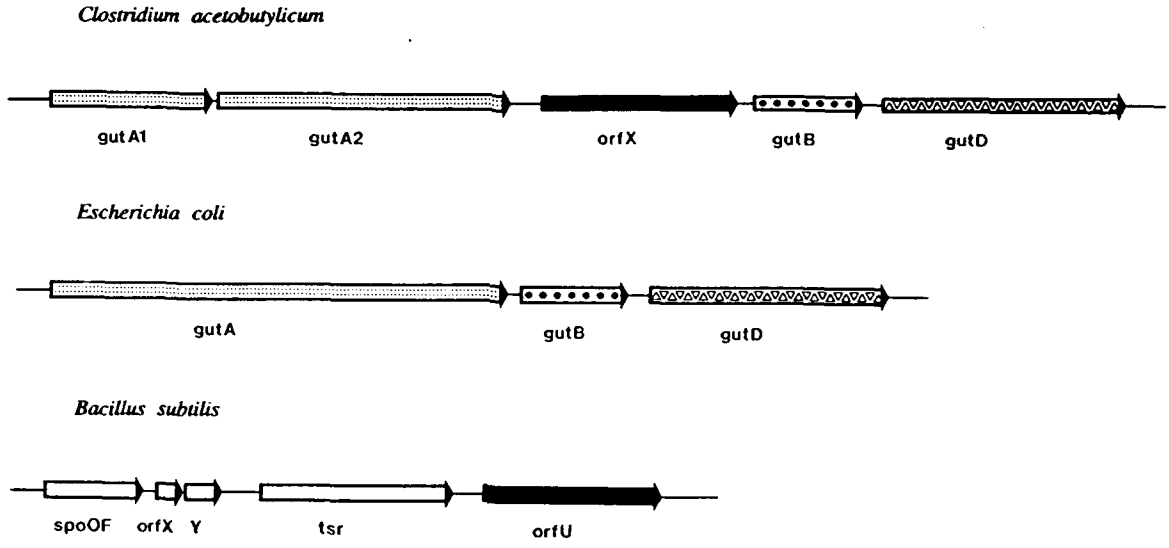


Fig. 17. *The arrangement of genes in the C. acetobutylicum gut operon, and the position of equivalent genes in E. coli and B.subtilis.* The encoded polypeptides of similarly shaded ORFs exhibit amino acid homology. The encoded enzymes are: *gutA*, PTS-II^{gut}; *gutB*, Enzyme III^{gut}; *gutD*, glucitol-6-P dehydrogenase, and; orfU & orfX (*C. acetobutylicum*), transaldolase. A sequence error in the illustrated *B. subtilis* region means that orfY and tsr form only 1 ORF, and encodes aldolase (J Cary, personal communication).

B. subtilis (Trach et al., 1988). Recently ORF U polypeptide has been shown to exhibit distant homology to yeast transaldolase (J Clary, personal communication), providing tentative evidence that the *C. acetobutylicum* ORF X gene product may be transaldolase (Fig. 17). The *gutD* gene (encoding glucitol dehydrogenase) seems an ideal target for integration as it is not normally required by the host, and presents an easy test for successful integration, ie., inability to grow on sorbitol as the carbon source.

Integrative vectors

Integrative vectors are ideally based on plasmids which are temperature sensitive for replication. Such a vector, containing cloned region of the host genome, may be introduced into the target cell and selected at a temperature permissive for replication. Successfully transformed cells may then be grown at the non-permissive temperature in the presence of the antibiotic to which the vector confers resistance. Under the these conditions plasmids are

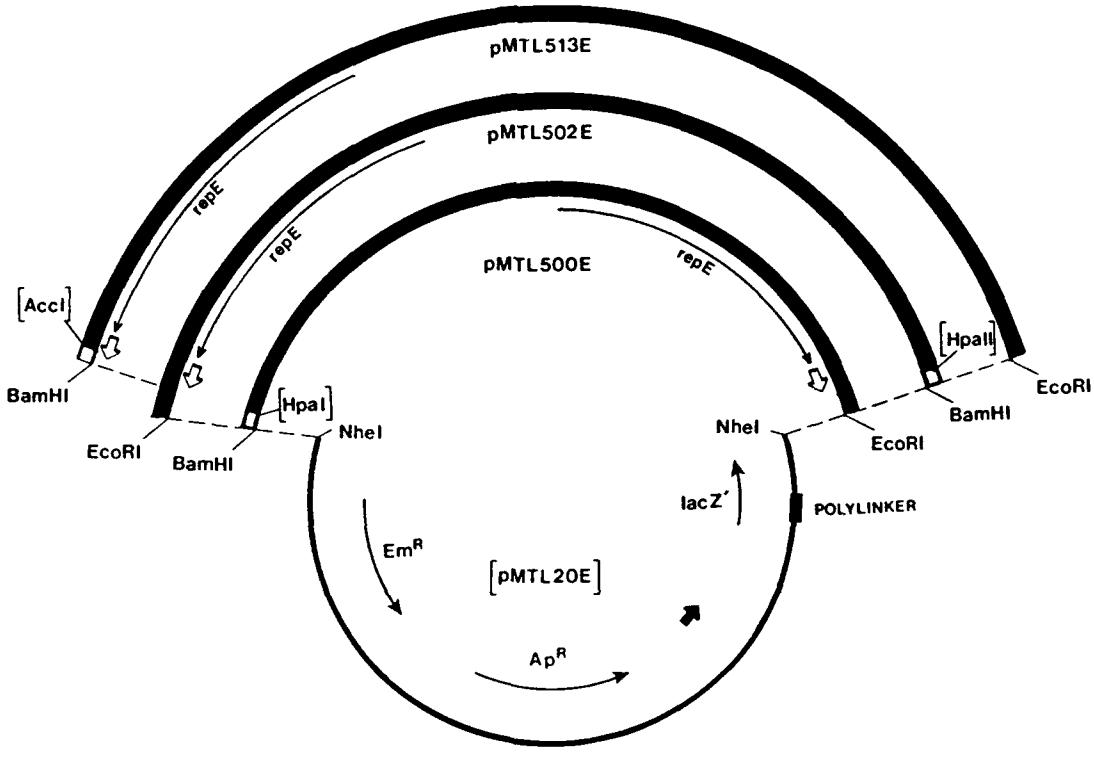


Fig. 18. ***Cloning vectors based on the pAMβ1 replicon.*** All plasmids were generated by insertion of the indicated pAMß1-derived DNA (see Swinfield et al., 1990) fragment (bold line) into the *Nhe*I site of pMTL20E (thin line). The *lacZ'* is therefore functional (blue colonies in the presence of XGal) unless inactivated by subsequent insertion of heterologous DNA into the polylinker region. Plasmid pMTL500E is a high copy number plasmid, while pMTL502E and pMTL513E have a low copy number. The general purpose cloning vectors pMTL500E and pMTL502E exhibit moderate segregational stability. Plasmid pMTL513E exhibits extreme instability in both *B. subtilis* and *C. acetobutylicum* (see Table 7).

rapidly lost from the population with the result that the only cells which can grow in the presence of the antibiotic are those in which chromosomal integration of the plasmid element occurs. With this in mind, attempts have been made to isolate a temperature-sensitive replication mutant of pMTL500E (Fig. 18) by *in vitro* mutagenesis. Plasmid DNA was incubated with hydroxylamine, as previously described (Minton, 1984), and the resultant damaged DNA used to transform *E. coli* cells to Ap^R. Total transformant colonies were then pooled (by flooding the agar plates with media), bulk plasmid DNA prepared and used to

transform *B. subtilis* to Em^R at 28°C. Colonies obtained were then replica plated onto fresh plates and grown for 24 h at 42°C. Approximately 10,000 *B. subtilis* colonies were screened in this manner, but only one putative ts mutant was isolated. Subsequent characterisation of this transformant, however, indicated that ts defect resided in the adenine methylase enzyme (*erm* gene).

In parallel to the above, the replication-impaired vector pMTL513E (Fig. 18) was examined as a possible integrative delivery system. This vector was derived by replacing the pAMß1 replication region of pMTL500E with the pAMß1 replicon of plasmid pMTL20Cß13 (Swinfield et al., 1990). Because this replicon contains a deletion which extends into the replication origin, the efficiency of replication is severely impaired. Thus, in the presence of the selective antibiotic *B. subtilis* cells carrying this plasmid exhibit a 4-fold increase in doubling time, while in the absence of selective pressure plasmid-free segregants arise at an extremely high frequency (Swinfield et al., 1990).

Use of pMTL513E to generate integrants formed by a single cross-over events

To investigate the potential of pMTL513E as a delivery sytsem, a 336 bp *Nhe*I-*Spe*I restriction fragment, internally located within the *gutD* structural gene was cloned into the polylinker of pMTL513E at its unique *Xba*I site. The plasmid obtained, pJEN2, was transformed into *C. acetobutylicum* NCIB 8052 and Em^R transformants selected. Interestingly,

PLASMID	% OF CELLS RESISTANT TO ERYTHROMYCIN	
	10 generations	20 generations
pMTL531E	99.5	99
pMTL500E	66	44
pJEN2 (pMTL513E)	0.4	0.01
CHR::pJEN2	96	92.3

Table 8. ***Segregational instability of pMTL513E (pJEN2) during growth of C. acetobutylicum in the absence of antibiotic selection.*** Cells were grown in 2 X YTG for 10 and 20 generations and the % of cells still Em^R estimated by deriving colony viable counts on media with and without Em. For comparative purposes, the results with pMTL500E and a stabilised derivative, pMTL531E (Swinfield et al., 1991), are shown. The Em^R phenotype of cells in which pJEN2 have integrated into the chromosome (CHR::pJEN2) exhibits a low level of instability (c. 0.4% per generation). All the resultant Em^S cells are also unable to grow on sorbitol.

pJEN2 transformed *C. acetobutylicum* at a significantly higher frequency (5-fold higher) than

the progenitor vector, pMTL513E, presumably as a result of carrying a homologous chromosomal DNA insert. A transformant containing the plasmid was then grown for 50 generations in the absence of antibiotic selection, before Em was added to the medium, the culture incubated for a further 8 hours and cells plated out on agar medium containing Em^R. As can be seen from Table 8, pJEN2 was rapidly lost from the cell population in the absence of selective pressure. Indeed, when the culture was plated out after 50 generations only 15 Em^R colonies were obtained. Using appropriate minimal agar plates, the cells from all 15 colonies were subsequently shown to be incapable of growth on sorbitol as the sole carbon source. Furthermore experiments indicated that loss of Em^R no longer occurred at the extremely high frequency initially observed in cells carrying autonomous pJEN2 (Table 8). Both observations strongly indicated that integration of pJEN2 had occurred at the *gutD* gene. A schematic representation of how pJEN2 could become inserted at the *gutD* locus of the *C. acetobutylicum* chromosome by Cambell integration is shown in Fig. 16. In this scheme a single recombinational cross-over results in duplication of the homologous *gutD* gene segment, concommitant with inactivation of the chromosomal copy. Double cross-overs would not inactivate the gene, nor result in a strain exhibiting segregational stabilisation of the Em^R determinant.

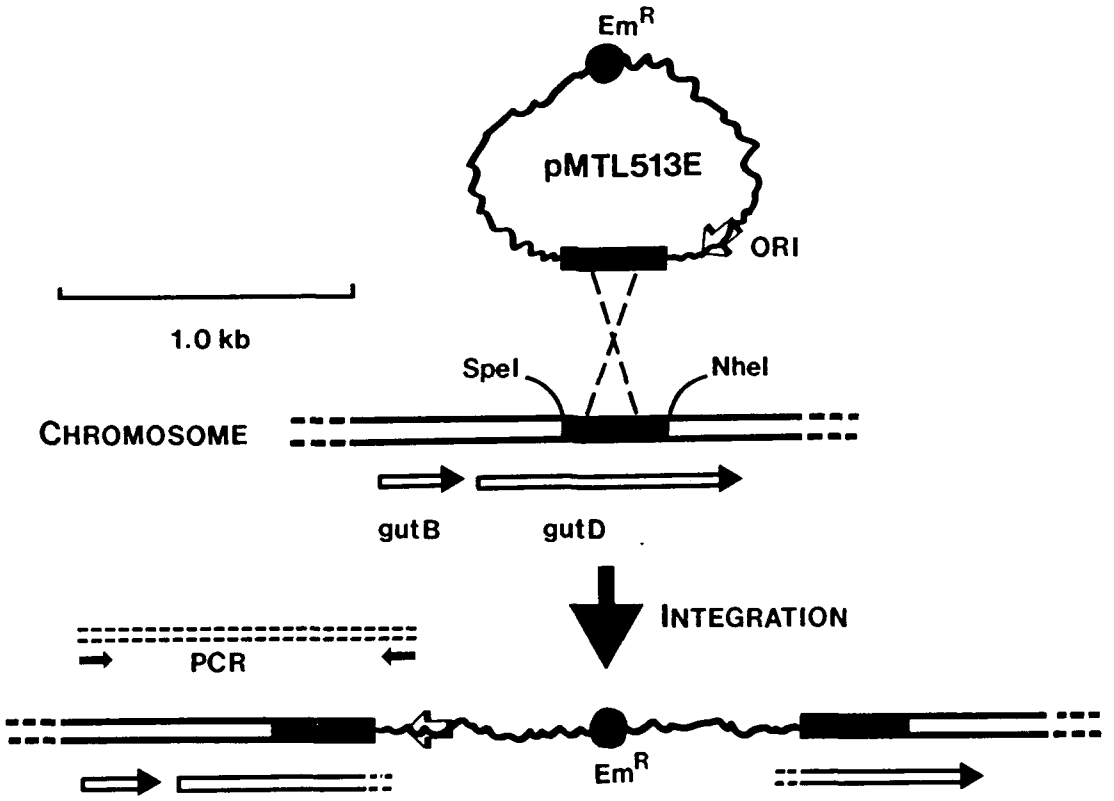


Fig. 19. *Schematic representation of Cambell-like integration of pMTL513E containing a gutD subfragment into the C. acetobutylicum chromosome.*

Confirmation that the stabilisation of Em^R segregation was a direct result of Campbell-like integration of the entire vector into the host chromosome at the *gutD* locus was obtained using PCR. Thus, oligonucleotide primers based on sequences within the chromosomally located *gutB* gene and the vector pMTL513E were shown to generate a DNA fragment of the expected size when employed in a PCR (Fig. 19). Although the Em^R phenotype of the pJEN2 integrant appeared, at a qualitative level, segregationally stable (see MIDTERM report), a more quantitative examination showed that in the absence of antibiotic selection significant numbers of cells were becoming Em^S at each generation, viz after 10 generations c. 4% of the cell population was Em^S (Table 8). That these cells represented cells in which pJEN2 had both excised from the chromosome and then been lost from the cell was confirmed by the fact that Em^S cells had regained the ability to grow on sorbitol as sole carbon source. A estimate of the actual rate of excision was made by screening for cells which had regained the ability to grow on sorbitol, but were still Em^R, ie., the plasmid excises from *gutD* and remains in the cell in an autonomous state. This showed that excision occurred in c. 0.04% of the cells at each generation.

Use of pMTL513E to generate integrants formed by a double cross-over events

The instability of the CHR::pJEN2 strain serves to highlight the unsuitable nature of integrants which arise by a single cross-over recombinational event. Stability may be ensured by selecting for integration of heterologous DNA by a double cross-over. The type of plasmid needed to achieve this should differ from pJEN2 in possessing a complete copy of *gutD* (pJEN2 contains only an internal region of the gene), into which heterologous genes are inserted. Such a vector will only generate SORB -ve integrants if a double cross-over occurs - single cross-overs, in which the whole plasmid integrates, will still be SORB +ve. Our eventual goal was a plasmid like pJEN2 in which the central portion of a cloned *gutD* gene is replaced by *lacI*. When this plasmid is introduced into *C. acetobutylicum* NCIB 8052, reciprocal recombination can take place between the 5' and 3' regions of *gutD* flanking *lacI*, resulting in the integration of *lacI* into the chromosome. However, such an event cannot be detected unless a selectable phenotypic trait is endowed upon the integrant. The solution is to link *lacI* to a selectable marker, which becomes co-integrated. As this cannot be *erm*, we elected to attempt to use *leuB*. Thus, cells would be transformed with selection for Em^R, a transformant grown in the absence of antibiotic for 50 generations in rich media, and cells plated on basal media lacking leucine. Integrants in which a double cross-over had occurred would then be selected on the basis of their Em^S and SORB -ve phenotype.

The first step in the construction of the desired plasmid was to insert a region of the *gut* operon into pMTL513E. Accordingly, a 1.1 kb *Bam*HI-*Sst*I fragment carrying the entire *gutD*

gene, and part of the upstream *gutB* gene was subcloned from pSORB20 into pMTL513E to give plasmid pSORB513 (Fig. 20). In parallel a second plasmid, pLACLEU (Fig. 20), was constructed in which a 1.45 kb *Nde*I-*Cla*I fragment carrying the *C.pasteurianum leuB* gene (Oultram et al., 1993) was co-cloned into pMTL23 along with a 1.3 kb *Eco*RI-*Pvu*I fragment carrying the *E. coli lacI* gene. The insert of this latter plasmid was subsequently excised as a 2.75 kb *Xba*I fragment and inserted into the *Spe*I site within the *gutD* gene of plasmid pSORB513. The plasmid obtained was designated pIB513 (Fig. 20). Prior to transformation of pIB513 into *C. acetobutylicum*, with the exception of *lacI*, all components were shown to be functional. Thus, the *leuB* moiety was shown to be functional by its ability to complement an

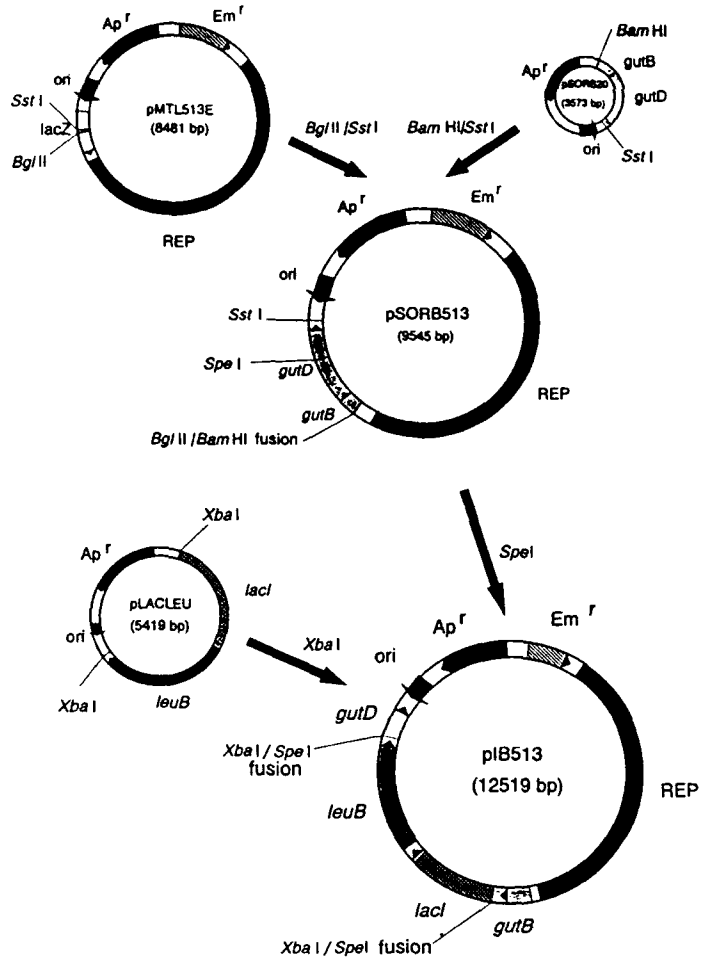


Fig. 20. ***Construction of the vector employed to target lacI to the NCIMB genome, plasmid pIB513.*** [see text for explaination]

appropriate mutant of *E. coli*. Similarly, the Gram-negative and Gram-positive selectable markers and replicons (Ap^R and ColE1, and Em^R and pAMß1, respectively) were shown to function in *E. coli* and *B. subtilis*. Thereafter, pIB513 was transformed into the Leu^- mutant of *C. acetobutylicum*, SA9, by electroporation and the Em^R transformants shown to be converted

to prototrophy. Lysates prepared from individual clostridial transformants were used to transform *E. coli* and these transformants shown to harbour a plasmid possessing a restriction pattern consistent with pIB513. Thus, pIB513 appears structurally stable in *C. acetobutylicum*.

A SA9 transformant carrying pIB513 was grown for 50 generations in the absence of antibiotic selection, and then plated out on both on minimal plates lacking leucine and broth plates supplemented with Em. An unexpectedly high number of colonies were obtained on both types of media (approx 10^3 per ml). A total of 100 leu^+ colonies were picked and shown, not unsurprisingly, to be still Em^R and capable of growth on Sorbitol as carbon source, ie., not integrants. Thus in this particular experiment, plasmid pIB513 was not lost at the expected rate, resulting in the selection of cells in which the plasmid still existed in the autonomous state. The experiment was therefore repeated twice more, except the de-selection and re-

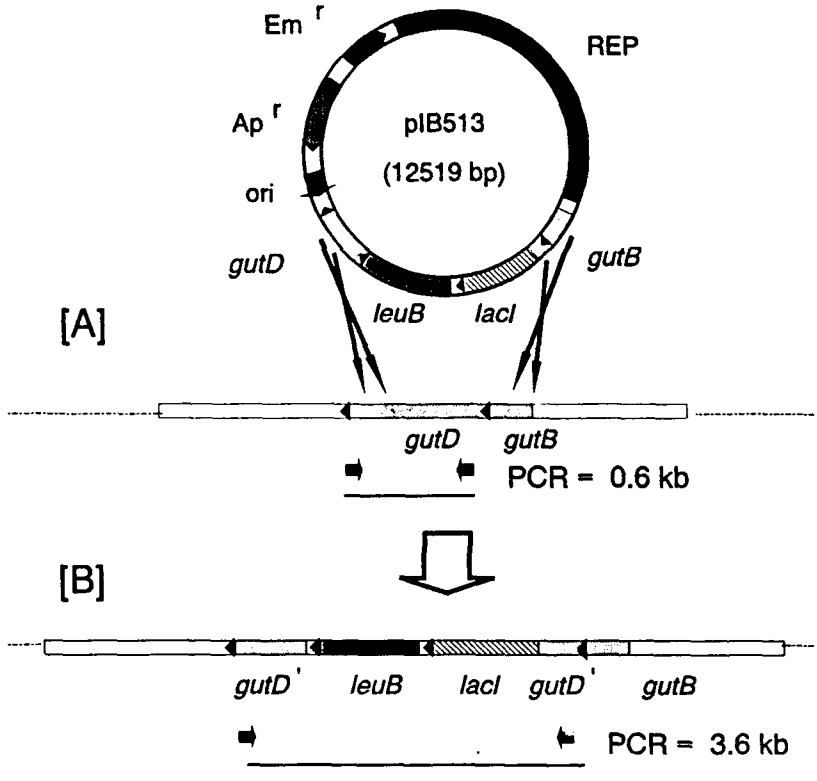


Fig. 21. *Gene replacement using pIB513.* Following two separate recombinantional evenets between homologous DNA on pIB513 and the NCIMB 8052 chromosome ([A]) the *gutD* gene of the latter is replaced by the copy on pIB513 containing *leuB* and *lacI* ([B]). Prior to gene replacement two opposing primers to the 5' and 3' ends of *gutD* amplify a 600 bp fragment in a PCR. Following integration of *leuB/ lacI* the amplified fragment increases in size to 3.6 kb.

selection steps were extended, such that the whole process took 2 weeks. From these two experiments 3 colonies were identified with the expected phenotype, viz., Leu^+, $Sorb^-$ and

Em^S. To characterise these "clones", use was made of a pair of oligonucleotide primers which in PCR amplify a 600 bp DNA fragment corresponding to the central portion of the *gutD* gene. The use of these primers in PCR with chromosomal DNA prepared from all 3 clones resulted in a c. 3.6 kb DNA fragment. This size exactly corresponds to that expected if the *gutD* gene contains the *leuB*::*lacI* insertion. Further evidence to support this contention was obtained by Southern blots (data not shown)

Failure of lacI to regulate fac

One of the three 3 *gutD*::*lacI*/*leuB* SBA9 integrants was chosen and transformed with a derivative of pMTL500F (pMTL500F*cat*) into which had been inserted a promoter-less copy of the staphylococcal CAT (chloramphenicol transacetylase) gene, inserted such that its transcription was under *fac* control. The levels of CAT attained in the resultant cells was, however, unaffected by the presence or absence of the gratuitous inducer, IPTG (data not shown). The reason for the apparent lack of repression/ induction were unclear, but may be due to a low level of production of LacI as a result both of low gene dosage, by virtue of a chromosomal location, or due to the *Bacillus* vegetative promoter transcribing *lacI* being inefficiently utilised by the clostridial RNA polymerase. One way to tackle the problem of low gene dosage would be to locate the *lacI* gene on the expression vector itself. Although the derivation of such a plasmid had previously proven unsuccessful, another attempt was made.

TIME[a] (hr)		CAT EXPRESSION LEVEL (% cell soluble protein) - IPTG	+ IPTG	MAGNITUDE OF INDUCTION (- fold)
	Exp. No.			
0	I	0.62	0.65	-
	II	0.58	0.89	-
	III	0.65	0.74	-
3	I	2.57	8.08	3.1
	II	1.82	9.17	5.0
	III	3.41	6.25	1.8
6	I	7.12	24.07	3.4
	II	10.11	14.44	1.4
	III	10.67	20.6	1.9
20	I	7.78	17.01	2.2
	II	9.32	21.85	2.3
	III	3.81	19,74	5.2

Table 9. *IPTG-mediated induction of cat expression in B. subtilis cells carrying plasmid pMTL500Flcat.* Cells were grown in L-broth to mid-logarithmic phase (A_{450} = 0.6), split in two and IPTG added (final concentration 400 μg/ ml) to one half of the culture. After 120 min of further growth cells were harvested, sonic extracts prepared, and assays for CAT activity undertaken

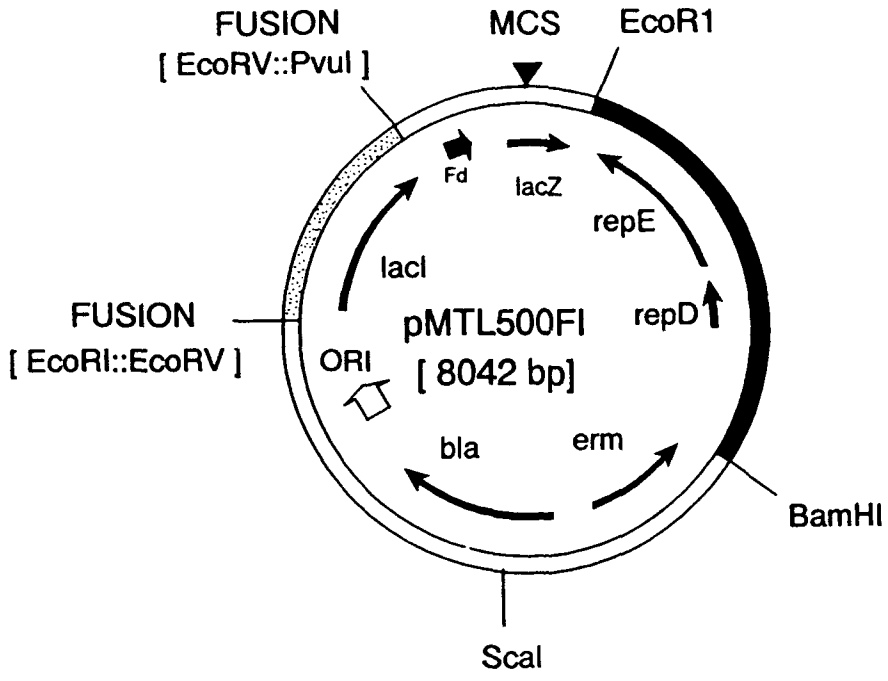


Fig. 22. ***The clostridial expression vector pMTL500FI.*** Plasmid pMTL500FI was derived from pMTL500F by inserting a 1.3 kb *Pvu*I-*Eco*RI (blunt-ended) fragment into the unique *Eco*RV site of pMTL500F (see Fig. 15).

A 1.3 kb *Pvu*I-*Eco*RI (blunt-ended) fragment carrying *lacI* (under the transcriptional control of the *B. subtilis veg*II promoter) was inserted into the *Eco*RV site of pMTL500F such that the gene is read in the same direction as *fac* and *bla*. The recombinant plasmid obtained, pMTL500FI (Fig. 22), in contrast to previous attempts, appeared as expected on the basis of restriction digests. Therefore, a 0.8 kb *MluI* fragment specifying a promoter-less copy of the pC194 *cat* gene, was inserted into the polylinker to give plasmid pMTL500FI*cat* and transformed into wild type cells of NCIMB 8052. Once again, no evidence of IPTG-mediated induction of *cat* expression was obtained. To investigate this further, plasmid pMTL500FI*cat* was transformed into *B. subtilis* and the experiments repeated. In this case the degree of induction was found to vary from between 2 to 5-fold, with induced cells producing up to 20% of their cells' soluble protein as CAT (see Table 9).

2.4 ATTEMPTED EXPRESSION OF BoNT/A H_C-ENCODING FRAGMENTS

2.4.1 Summary

In the absence of a regulated system, attempts were made to effect the constitutive expression of *botA* subfragments from the *fac* promoter. To aid in the subsequent purification of the recombinant polypeptides produced, a strategy was formulated whereby they would be produced as a fusion protein with glutathione-S-transferase (GST), whose encoding gene exhibits a similar codon usage to clostridial genes. To accomplish this, DNA encoding the H_C fragment of BoNT/A (aa 855 to 1296) was fused to the extreme 3'-end of the GST gene, using PCR methodologies. To ensure eventual translation of the transcribed gene fusion in a Gram-positive host, a synthetic sequence specifying the ribosome binding site (RBS) of the TeTx gene was positioned immediately 5' to the translational start codon of the GST gene. The completed gene fusion was placed under *fac* transcriptional control by its insertion into pMTL500F. No evidence for the production of a recombinant protein was obtained when Western blots were performed on the lysates of *E. coli* cells carrying the resultant plasmid, pGAC501F, using either anti-BoNT/A or anti-GST antibody. Although cells carrying pGAC501F produced abnormal amorphous growth on solidified media, no evidence for the presence of inclusion bodies was forthcoming. Plasmid pGAC501F was subsequently found to be incapable of transforming either *B. subtilis* or *C. acetobutylicum*, a consequence, it is believed, of the production of the desired fusion protein. Derivative plasmids of pGAC501F were constructed in which the region encoding the entire BoNT/A H_C fragment was replaced with *botA* DNA encoding the NH_2- or COOH-terminal half of the H_C fragment (plasmids pGAC503F & pGAC504F, respectively). These new plasmids were now able to transform both Gram-positive hosts. The presence of a novel fusion protein could not, however, be detected in the lysates of transformed cells. Preliminary experiments, involving placement of the Fd RBS immediately 5' to the GST start codon, suggest that the TeTx RBS may be responsible for the lack of detectable protein.

2.4.2 Results and Discussion

Construction of BoNT/A H_C ::Glutathione-S-transferase gene fusions

In view of the difficulty encountered in attempting to regulate the *fac* promoter, a decision was made to push ahead with constitutive expression of BoNT gene subfragments in *C. acetobutylicum* using pMTL500F. By this stage of the project studies being undertaken by

Division of Toxinocology staff at USAMRID had shown that a recombinant polypeptide corresponding to a BoNT/A H_C fragment (equivalent to tetanus toxin "C" fragment) was shown to be protective in mice. Furthermore, these same studies had demonstrated that

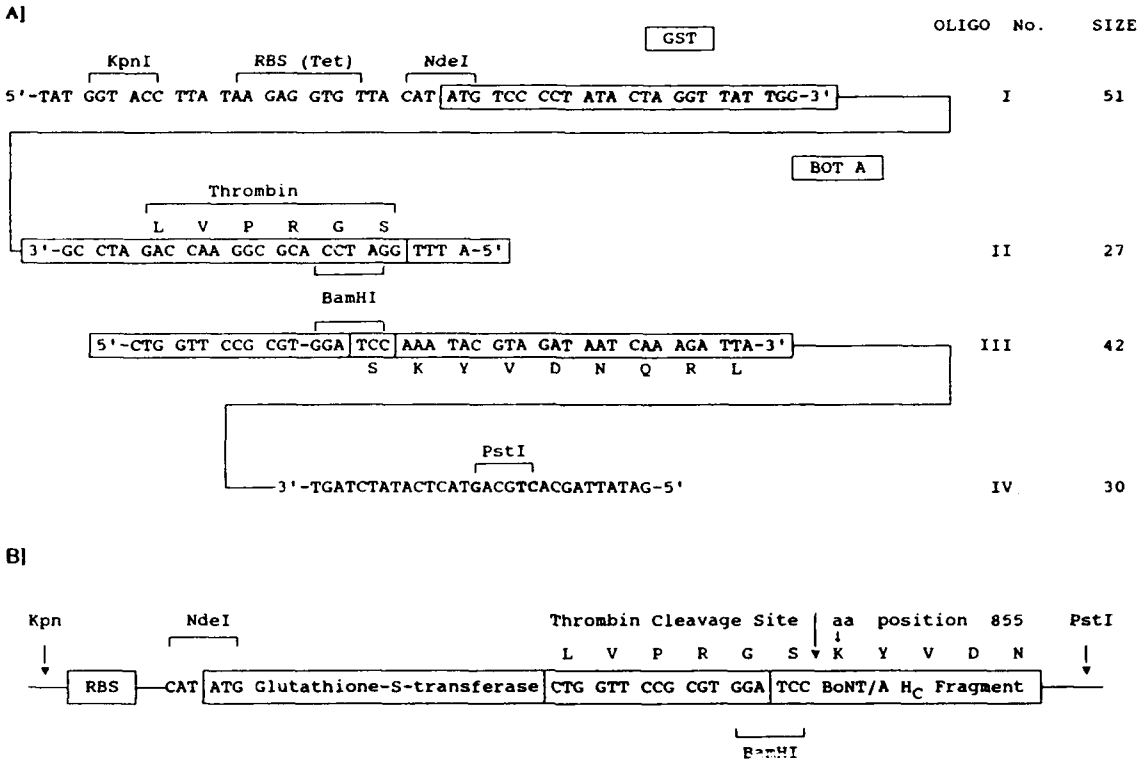


Figure 23. ***Strategy for the construction of a GST::BoNT/A H_C fusion protein by PCR.***

A] The two oligonucleotides I and II are used in PCR to amplify a 690 bp fragment from pGEX-2T encoding the entire Glutathione-S-transferase (GST) protein. The 5' tail of oligo I will specify, in addition to the 5' end of the GST structural gene, the ribosome binding site of the tetanus toxin gene, flanked by restriction sites for *Kpn*I and *Nde*I. Oligo II will essentially encode the Thrombin site of plasmid pGEX-2T, with a small 3' tail of complimentarity to the *botA* gene. The two oligonucleotides III and IV are employed to amplify a 1.45 kb fragment of the botA gene. Oligo III, in addition to specifying 9 amino acids from the NH_2-terminus of the BoNT/A H_C fragment (essentially beginning with the Ser residue at position 854 of BoNT/A), has a 5' tail complimentary to the Thrombin site of pGEX-2T. Oligo IV is complementary to a sequence some 100 bp downstream of the *botA* translational stop codon, and contains the necessary mismatches to allow the creation of a *Pst*I site. Oligos II and III have been designed such that the DNA fragments amplified in the respective I+II and III+IV PCR's will carry an identical 21 bp sequence at their 3' and 5' ends, respectively. This overlap may be used in a subsequent PCR to join the two fragments, giving the desired GST::BoNT fusion, illustrated in *B]*. The presence of the *Kpn*I and *Pst*I sites allow the insertion of this fragment into the polylinker of pMTL500F/ FI.

recombinant production was only achieved in the form of a fusion protein, a consequence of the genetic fusion of the appropriate *botA* subfragment with the *malE* (maltose binding protein) gene of *E. coli*. Based on these findings it was decided to limit expression studies to the H_C fragment of the neurotoxin and to produce the toxin as a fusion protein. In our case, however, we chose to use the glutathione-S-transferase (GST) gene of the Pharmacia plasmid pGEX-2T (Smith and Johnson, 1988) rather than the *malE* gene. As with MalE, the fusion protein produced can be purified by affinity chromatography, the BoNT moiety being recovered

following cleavage with thrombin. The glutathione-S-transferase gene was considered more appropriate than *malE*, however, as it has an A+T content of 62%, resulting in a codon usage near to that found in clostridia.

To effect the fusion of GST and BoNT/A encoding regions, the strategy outlined in Fig 23 was formulated. Accordingly, primers I & II were used in PCR to amplify a 0.6 kb fragment specifying GST, and primer pair III + IV was used to amplify the BoNT/A H_C-encoding fragment. These fragments were gel isolated, pooled and used in a subsequent PCR employing primer pairs I + IV. Inexplicably, no DNA product was obtained. Therefore, the two fragments were independently cloned into pCR1000. The two inserts were subsequently excised as a *Kpn*I/*Bam*HI (GST) and a *Bam*HI/*Pst*I (BoNT/A) fragment, pooled and ligated to *Kpn*I/*Pst*I cleaved pMTL21, and a plasmid selected (pGAC1) in which the two fragments were co-inserted. At this stage further experiments had shown that a GST::BoNT/A H_C fusion could be generated, simply by mixing the two plasmids pGEX-2T and pCBA3, and undertaking a PCR with all 4 primers (I, II, III & IV) present. Samples taken from such a reaction were shown to contain 3 DNA bands, corresponding in size to that encoding GST, BoNT/A and a GST::BoNT/A fusion. The latter band was subsequently cloned directly into pCR1000. The clones obtained were not, however, processed any further as by this time a fusion of the two "genes" had been derived by standard cloning procedures.

Prior to the generation of pGAC1, the entire nucleotide sequences of the component fragments were determined to check for PCR-induced errors. None were found. Once pGAC1 was obtained, the junction between the GST- and BoNT/A H_C-encoding fragment was also authenticated by sequencing. Interestingly, during the initial cloning of the PCR products of primers III + IV, a clone was obtained which had a deletion at the 3'-end of the BoNT/A H_C-encoding region. In essence, 39 amino acids were deleted from the COOH-terminus. This variant was also fused to the GST encoding fragment, in pMTL21, to give pGAC2. The inserts of both plasmids were subsequently excised and sub-cloned into pMTL500F. The plasmids obtained were designated pGAC501F and pGAC502F, respectively. As a control, the two inserts were also cloned into pMTL500E, yielding pGAC501E and pGAC502E, respectively. Noticeably, all the clones obtained exhibited abnormal growth on solidified media. When streaked onto agar the growth that developed was sparse and the colonies had a amorphous atypical appearance. No evidence for the presence of inclusion bodies could be obtained by phase contrast microscopy. Lysates from each type of clone were examined by SDS PAGE, but no additional polypeptide bands of the expected size were evident. Similarly no bands were detected in a Western blot using anti-BoNT/A polysera.

Attempted transfer of pGAC501F & pGAC502F to C. acetobutylicum

Having constructed the four different recombinant plasmids attempts were made to transform them into *C. acetobutylicum* using electroporation. However, although transformants of pGAC501E and pGAC502E were obtained (ie., those plasmids derived from pMTL500E in which expression of gene subfragments will not occur in a Gram-positive

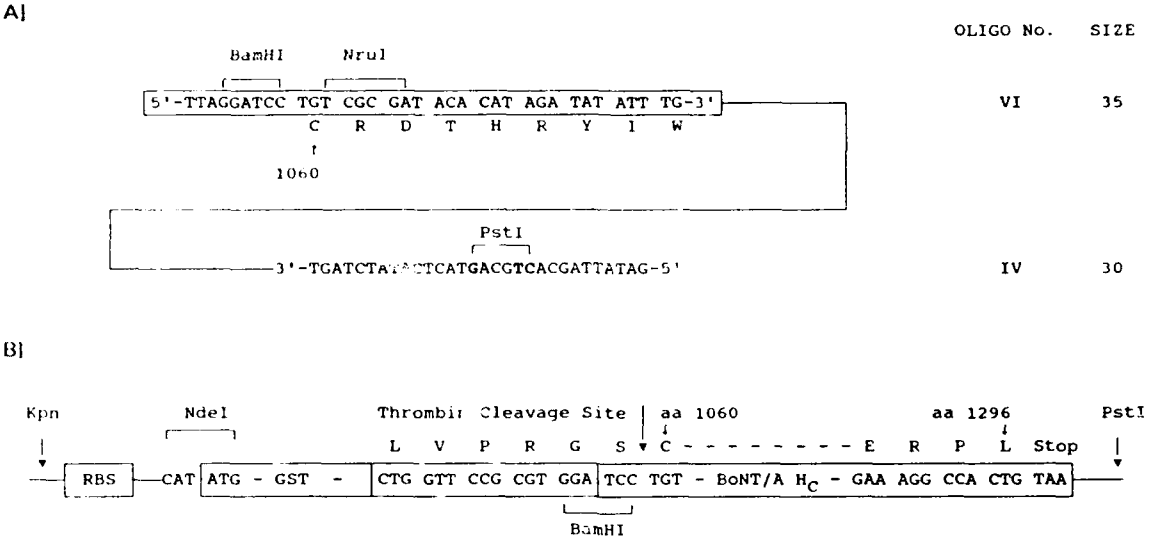


Figure 24. ***Strategy for the construction of a fusion protein between the COOH-terminus of BoNT/A H_C and GST by PCR.***

A] A new oligonucleotide, VI, was used in conjunction with the oligonucleotide IV to amplify a 0.7 kb fragment encoding the COOH-terminal domain (aa 1060 to 1296) of BoNT/A H_C. The amplified fragment was then ligated to the previously amplified 690 bp fragment encoding GST (see Fig. 23), using their complementary *Bam*HI sites, to generate the fragment illustrated in *B]*. The presence of the *Kpn*I and *Pst*I sites allowed the subsequent insertion of this fragment into the polylinker of pMTL500F.

bacterium), no transformants were obtained with any plasmid derived from pMTL500F. As DNA passaged through *Bacillus subtilis* has been found to transform *C. acetobutylicum* with higher efficiencies, attempts were made to transform *B. subtilis* 34.1 (*spo trpC*). Once again no transformants were obtained.

The two possible explanations for this lack of transformation would appear to be that either:- (i) that the replicon of the pMTL500F derived plasmids has in some way been disabled during the construction of the recombinant pGAC plasmids in *E. coli*, or; (ii) expression of a GST::BoNT/A fusion in *C. acetobutylicum* is lethal. With regard to the first possibility, a comprehensive series of digests with various endonucleases, however, has failed to yield any restriction patterns that disagree with that predicted. Any deletion/ re-arrangement would therefore have to be very minor indeed. Subsequently, a deletion variant of pGAC501F was

constructed by deleting the DNA specifying the GST::BoNT/A fusion protein. This plasmid was shown to be able to transform both *B. subtilis* and *C. acetobutylicum*, demonstrating that the replicative moiety of pGAC501F remained functional. These results strongly suggested that expression of the GST::BoNT/A encoding region is detrimental to the cell.

Construction of further BoNT/A -encoding derivatives of pMTL500F

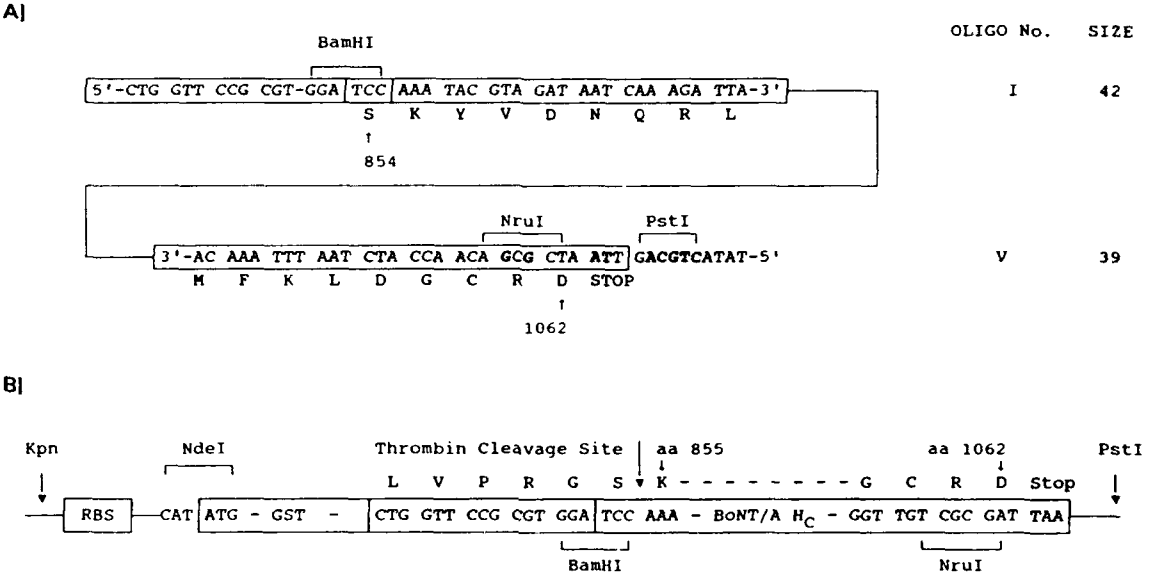


Figure 25. ***Strategy for the construction of a fusion protein between the NH_2-terminus of BoNT/A H_C and GST by PCR.***

A] A new oligonucleotide, V, was used in conjunction with the oligonucleotide III to amplify a 0.65 kb fragment encoding the NH_2-terminal domain (aa 854 to 1062) of BoNT/A H_C. The amplified fragment was then ligated to the previously amplified 690 bp fragment encoding GST (see Fig. 23), using their complementary *Bam*HI sites, to generate the fragment illustrated in *B]*. The presence of the *Kpn*I and *Pst*I sites allowed the subsequent insertion of this fragment into the polylinker of pMTL500F.

To clarify the matter with regard to the apparent toxicity of the fusion protein encoded by pGAC501F, a number of new plasmid derivatives were constructed. Initially equivalent plasmids to pGAC501F were constructed but in which only half of the BoNT/A H_C -encoding region was fused to GST. In the one case an approx. 0.7 kb fragment encoding the COOH-terminal portion of BoNT/A H_C (amino acids 1060 to 1296) was generated in PCR using primers VI and IV (see Fig. 24). In another case the NH_2-terminal portion of BoNT/A H_C was PCR amplified using primers III and V (Fig 25). Both primers were fused to the same GST-encoding fragment as was present in pGAC501F by virtue of a created *Bam*HI site. In addition to these plasmids, two further derivatives were also constructed. Plasmid pGAC505F was constructed in which the the BoNT/A H_C-encoding region was fused directly to the first few codons of the *lacZ'* gene. This was achieved by inserting the BoNT/A H_C -encoding, 1.45 kb *Bam*HI-*Pst*I fragment of pGAC501F directly between the *Bgl*II and *Pst*I sites of

pMTL500F, ie., in the absence of the GST gene. Finally, as a control, a plasmid was constructed, pGAC506F, which contained GST-encoding DNA alone. This was derived by simply inserting the 690 bp *Kpn*I/ *Bam*HI fragment amplified by oligonucleotides I and II (see Fig. 23) directly between the *Kpn*I and *Bgl*II sites of the polylinker region of pMTL500F. A schematic representation of all derived plasmids is given in Fig. 26.

In contrast to the two previous plasmids, all four new plasmids (pGAC503F, pGASC504F, pGAC505F & pGAC506F) could be transformed into both *B. subtilis* and *C. acetobutylicum*. Although no gross differences in colony morphology was evident between cells containing the four plasmids in either Gram-positive host, *E. coli* cells carrying pGAC503F (ie., the NH_2-terminus of BoNT/A H_C) gave atypical colonies on agar media. Cultures of all three bacterial

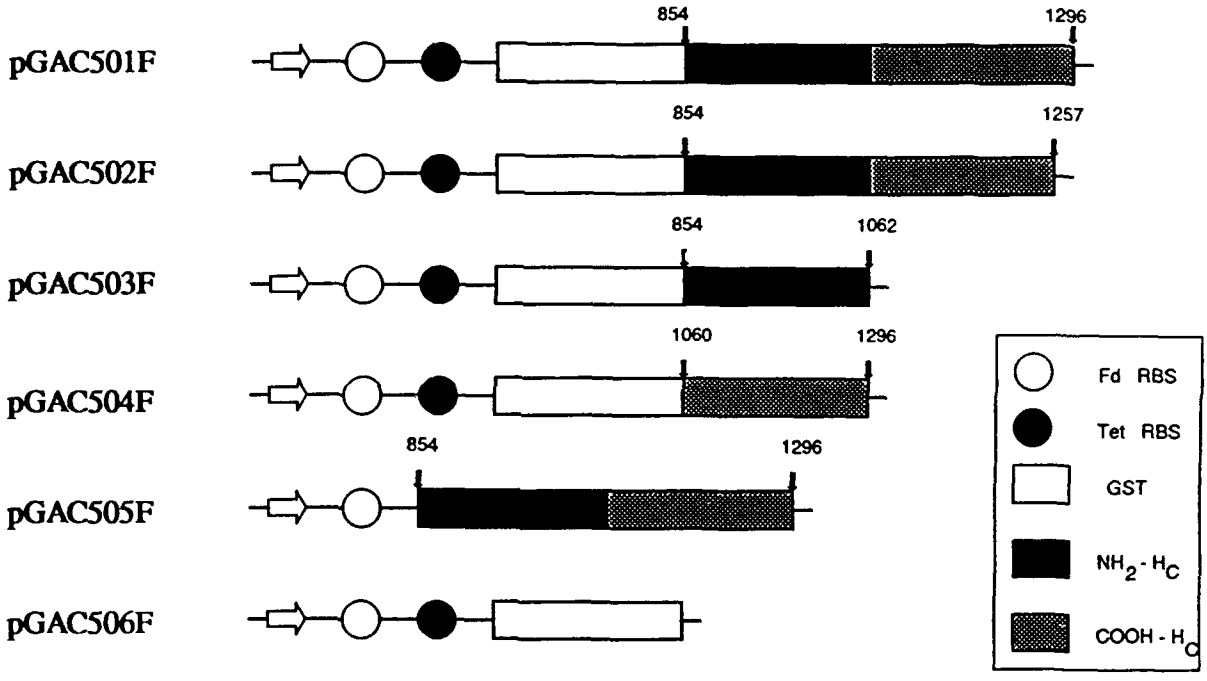


Figure 26. ***Plasmids based on pMTL500F used in attempts to obtain expression of botA subfragments.*** The components are as indicated in the right hand box. The open arrow corresponds to the Fd promoter. The numbers above each map indicate the amino acid number (relative to the complete toxin) at which the BoNT/A-derived regions begin and end.

hosts, carrying all 4 new plasmids were grown up overnight and cell lysates prepared. These were subjected to SDS-PAGE, and comassie-stained gels examined for the presence of novel protein bands. None were detected. The electrophoretograms were therefore subjected to Western blots and probed both with BoNT/A antisera and GST antisera. Purified BoNT/A was used as a control for the former, and an *E. coli* lysate derived from cells carrying the plasmid pGEX-2T was used as a control for the GST antisera. With the BoNT/A antisera, no novel protein bands were evident in any of the lysates tested. In contrast, with GST antisera, a band corresponding in size to that of GST was present in the lysates derived from both *B.*

subtilis and *E. coli*, but not *C. acetobutylicum*. The intensity of the "signal" was, however, orders of magnitude lower than that obtained in a lysate of *E. coli* cells carrying the control plasmid pGEX-2T.

2.5 STATUS OF EXPRESSION STUDIES ON TERMINATION OF THE CONTRACT

In the closing stages of the project we began to suspect that our inability to detect expression of BoNT/A fusion proteins could be attributable to the RBS sequence we had placed immediately 5' to the GST-encoding moiety, based on that of the TeTx gene. In a parallel piece of work, we had inserted an *E. coli*-derived gene into pMTL500F such that its RBS was replaced by that of Fd. The level of expression, in all three hosts tested (*E. coli*, *B. subtilis* and *C. acetobutylicum*) was such that the encoded recombinant protein attained levels representing approx. 9% of the cells soluble protein. This clearly showed that the Fd RBS could be efficiently utilised in both Gram-negative and Gram-positive hosts. The apparent lack of detectable protein in cells harbouring pGAC505F (in which the *lacZ*'::BoNT/A fusion is effectively coupled to the Fd RBS) was, however, not consistent with this notion. On re-examination of the procedure used to derive pGAC505F, however, it was found that insertion of the BoNT/A-encoding 1.45 kb *Bam*HI-*Pst*I fragment between the *Bgl*II and *Pst*I sites of pMTL500F does not result in an "in-frame" fusion of the *lacZ*' and *botA* coding regions. The mistaken assumption that in-frame fusion would occur was caused by a "rogue" computer printout of the pMTL500F sequence in which a nucleotide base from within the polylinker region was missing. On paper, the two coding regions could be simply converted to the same reading frame by, cleaving pGAC505F with *Xba*I, blunt-ending with Klenow polymerase, and then self-ligating. This modification is currently being undertaken. The efficiency with which blunt-ended, *Xba*I-cleaved pGAC505F DNA self-ligates, however, is proving to be extremely low. Amongst other explanations, this could be because cells transformed with a pGAC505F derivative in which LacZ'::BoNT/A is produced are not viable. Despite this, we have obtained several clones in which the plasmid has lost the *Xba*I site, but the junction between *lacZ* and *botA* has yet to be sequenced. In the meantime, a more effective way of utilising the Fd RBS has been devised.

In all plasmids carrying the GST gene, the TeTx RBS is flanked by *Nde*I restriction sites. To directly compare the relative efficiency of the TeTx RBS and that of Fd, DNA carrying the former was deleted from plasmid pGAC506F by its cleavage with *Nde*I and subsequent self-ligation. Lysates have been prepared from *E. coli* cells carrying the plasmid obtained (pGAC516F), it's progenitor (pGAC506F) and pGEX-2T, and subjected to SDS PAGE. The use of GST antisera in preliminary Western blots of the resultant gels appeared to indicate that the production of GST in cells carrying pGAC516F is significantly higher than in cells

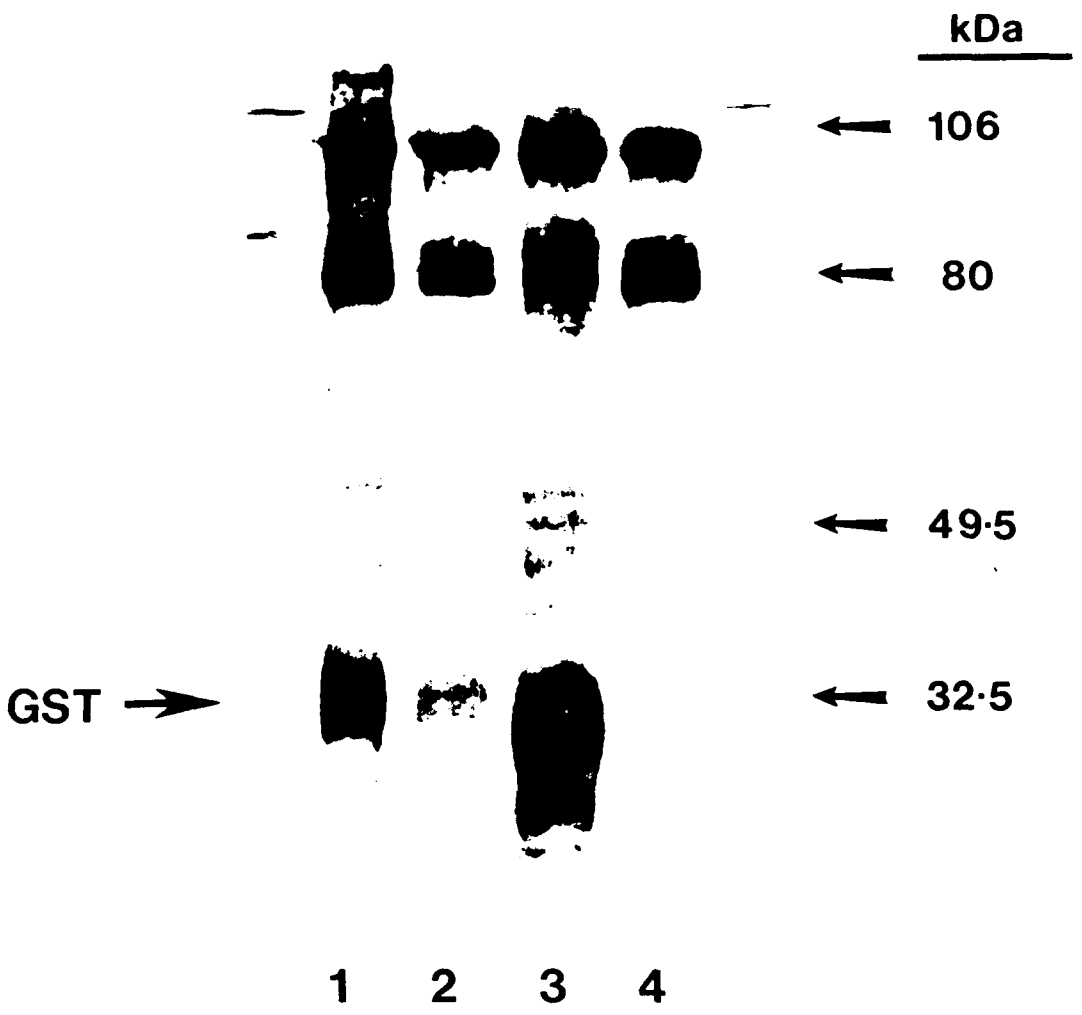


Figure 27. ***Western blots of E. coli lysates carrying pGAC506F, pGAC516F and pGEX-2T.*** Samples were prepared as described in materials and methods. Following SDS PAGE, the electrophoretograms where blotted with anti-GST antisera. Lanes: 1, *E. coli* [pGAC516F]; 2, *E. coli* [pGAC506F]; 3, *E. coli* [pGEX2T], and; 4, *E. coli* [plasmid-free].

harbouring pGAC506F (see Fig. 27). Plasmid pGAC516F has now been transformed into *B. subtilis* and *C. acetobutylicum*, and estimates of the level of recombinant GST produced are about to be undertaken. Should the results of these experiments prove encouraging, then similar *Nde*I deletions can be made to those plasmids encoding BoNT/A-derived polypetides, pGAC501-504F.

CONCLUSIONS

[1] CLONING OF *Clostridium botulinum* NEUROTOXIN GENES

A major target of this contract was to derive the entire nucleotide sequences of the *C. botulinum* genes encoding type B, E, F and G neurotoxin. This objective has been succesfully accomplished, and the complete nucleotide sequences of the BoNT genes of the *C. botulinum* strains Danish (type B), NCTC 11219 (type E), Langeland (type F) and 89G (type G) have now been determined. As a result, taken together with our previously determined type A gene sequence and sequences determined by other laboratories, a complete amino acid sequence of a representative toxin from all 7 serotypes is now available. Comparative analysis of this catalogue of sequences should considerably facilitate future studies concerned with structure/ function and vaccine development.

The data derived has shown that BoNT/B, BoNT/E, BoNT/F and BoNT/G are composed of 1291, 1252, 1278 and 1297 amino acids (aa), respectively, making the type E serotype the smallest characterised BoNT. Comparative alignment of translated aa sequences, and BoNT/A, C, D, and TeTx, demonstrates that clostridial neurotoxins are composed of highly conserved aa domains interspersed with aa tracts exhibiting little overall similarity. On the basis of aa similarity, TeTx is indistinguishable from a BoNT. In total 63 aa, out of an average 440, are absolutely conserved between L chains. and 93 out of 842 between H chains. The most divergent region corresponds to the carboxyterminus of each toxin, reflecting differences in specificity of binding to neurone acceptor sites. The relative order of relatedness varies according to which dichain component is compared. Recombinational events between different *bot* genes may therefore have taken place during evolution.

The BoNT/E and BoNT/B of this study show only minor differences to those of other strains. Conversely, the amino acid sequence of the BoNT/F determined in this study (isolated from a proteolytic *C. botulinum*, Langeland) exhibits considerable divergence from that of a BoNT/F derived from a non-proteolytic strain of *C. botulinum* (ATCC 23387), and the BoNT/F produced by a strain of *C. baratii* (ATCC 43756). Thus, the L- and H-chain of Langeland and ATCC 43756 share only 63% and 79%, respectively. The degree of homology shared by their L-chains is equivalent to that seen between the L-chains of BoNT/B and BoNT/G (61%). Data obtained in this laboratory during the course of the development of DNA probes has indicated that the degree of divergence between the neurotoxins of proteolytic and non-proteolytic type B *C. botulinum* strains may mirror that seen between type

F strains. Indeed, a recently published (Campbell et al., 1993) partial sequence (361 amino acids) of the H-chain of a BoNT/B produced by a non-proteolytic *C. botulinum* type B strain, exhibits 96% identity with the equivalent region of the BoNT/B of this study.

Divergence between toxins of a single serotype can have serious implications for any strategy in which a polypeptide subfragment of a toxin is being proposed as a subunit vaccine. Thus, for instance, the H_C fragment of the BoNT/F produced by one *Clostridium spp.* may not elicit protection against the BoNT/F produced by a second *Clostridium spp.* At present, however, the extent of divergence within the BoNT gene pool is unclear. An appreciation of the magnitude of this potential problem could be obtained by undertaking a survey of DNA variation in all available strains, employing a simple PCR screening procedure. This would involve preparing rapid small-scale genomic preparations from each strain, using serotype-specific primers to amplify a selected region of the BoNT structural gene, subjecting the amplified product to digestion with selected restriction enzymes and then comparing the fragment patterns generated using agarose gel electrophoresis. It's feasibility is demonstrated by the data in Table 10 & 11. In Table 10 the indicated fragments are those that would be generated if the PCR-amplified L-chain encoding region of the BoNT/E gene of *C. botulinum*

*Alu*I		*Dde*I		*Dra*I		*Mae*3		*Mbo*I		*Mnl*I	
BotE	ButE	BotE	ButE	BotE	ButE	BotE	ButE	BotE	ButE	BotE	ButE
328	328	499	499	**528**	-	717	717	654	654	-	**726**
285	285	410	410	251	251	**291**	-	**505**	-	**435**	-
268	268	298	298	205	205	258	258	-	**307**	387	387
216	216			-	**197**	-	**185**	-	**198**	**291**	-
169	169			155	155	-	**106**	77	77		
				94	94						

Table 10. *Predicted restriction patterns of the PCR-amplified L-chain encoding regions of the BoNT/E genes of the C. botulinum strain NCTC 11219 (BotE) and the C. butyricum strain ATCC 43181 (butE). The size of fragments is given in bp. Fragments unique to a gene are emboldened.*

NCTC 11219 and *C. butyricum* were digested with the indicated enzymes. While certain enzymes will generate identical patterns with the fragments amplified from the two genes (eg., *Alu*I and *Dde*I), a substantial number (nearly half of all those enzymes predicted to have at least 3 recognition sites in L-chain encoding DNA) will give discernible differences in restriction patterns, eg., *Mbo*I and *Mnl*I. As shown in Table 11, an identical situation is encountered if one undertakes the same analysis with the L-chain encoding regions of the BoNT/F gene of strain Langeland and the BoNT/F gene of the non-proteolytic strain ATCC `23387. The two BoNT/E genes differ by 27 nucleotides out of a total of 1266, while the two BoNT/F genes possess 33 dissimilar nucleotides out of a total of 1313. It can therefore be seen that this method is capable of a high degree of sensitivity with regard to the detection of nucleotide divergence. Suspected divergence and existence of distinct sub-populations could

AluI		*DdeI*		*DraI*		*Mae3*		*MboI*		*MnlI*	
ProF	NonF	ProF	NonF	ProF	NonF	ProF	NonF	ProF	NonF	ProF	NonF
570	570	**761**	-	561	561	**590**	-	614	614	615	615
-	**421**	-	450	263	263	-	**337**	**343**	--	228	228
198	198	-	**311**	256	256	320	320	-	**303**	155	155
157	-	248	248	189	189	-	**253**	162	177	128	128
106	106	164	164			**186**	-	95	95		
		140	140			147	147				
						-	**109**				

Table 11. *Predicted restriction patterns of the PCR-amplified L-chain encoding regions of the BoNT/E genes of the C. botulinum group I proteolytic strain Langeland (ProF) and the non-proteolytic group II strain 202F (NonF). The size of fragments (in bp) unique to a gene are emboldened.*

then be confirmed by direct nucleotide sequencing of the amplified regions using appropriate primers.

[2] EXPRESSION SYSTEM DEVELOPMENT

The second major objective of this contract was to develop a clostridial expression system and use it to express BoNT gene subfragments. Attempts to elicit the transfer of plasmid DNA vectors into 20 different strains of the intended host, *Clostridium sporogenes*, by either electro-transformation of by conjugative mobilisation, however met with no success. Rather than persevere with this *Clostridium sp.*, the genetically amenable species *Clostridium acetobutylicum* NCIB 8052 was chosen as an alternative host for the proposed expression work. Efforts initially focused on imposing regulatory control on the *fac* promoter system by seeking to obtain expression of *lacI* in *C. acetobutylicum*. Although this gene was succesfully introduced into *C. acetobutylicum*, both by integrating it into the chromosome and by incorporating it into the backbone of the expression vector employed, *fac* remained un-regulated. This was attributed to inefficient expression of the *lacI* gene. Thereafter, attempts were made to constitutively express gene fusions between H_C-encoding regions of the *botA* gene and the gene encoding glutathione-S-transferase (GST), by placing appropriate DNA downstream of the *fac* promoter. At the time of writing no positive evidence that a fusion protein is being produced in *C. acetobutylicum* has been obtained.

Although progress with this aspect of the study has been somewhat disappointing, it is probably an accurate reflection of the inherent difficulties one would expect to encounter, compared to *E. coli*, when attempting to genetically manipulate a clostridial species. Even so, the situation upon termination of the contract is at a hopeful stage. It would appear to have been an unfortunate decision to opt for the RBS of the clostridial TeTx gene rather than that of the clostridial ferredoxin gene. At the time, however, there was no practically derived

evidence to suggest that one was any better than the other. We presently beleive that production of a GST::BoNT/A H_C fusion protein in the two Gram-positive bacteria tested (*B. subtilis* and *C. acetobutylicum*) is lethal to the cell. Whether this lethality is an intrinsic property of BoNT/A itself or a consequence of its fusion to GST, remains an open question. With hindsight, it may have been preferable to fuse BoNT-encoding DNA to a gene encoding a secreted protein. The fusion product would then be exported into the culture medium. A number of genes encoding clostridial genes whose products are secreted have been cloned and sequenced. In contrast to GST-based polypeptides, the characteristics of such fusion proteins would, however, not facilitate their subsequent purification. Although, the use of the secreted MalE protein of *E. coli* would have circumvented problems of purification, its use is prejudiced by the inappropriate codon usage of its gene, and the fact that its RBS and signal peptide sequence are unlikely to function in a Gram-positive. Replacement of the latter two elements with the equivalent of a clostridial gene (eg., *celA*) would obviously circumvent these barriers to expression, but introduce an additional level of complexity to the system. A more realistic approach would be to fuse the BoNT-encoding DNA to the staphylococcal protein A gene, which is itself of Gram-positive origin.

In conclusion, the termination date of this contract arrived too soon for the potential of clostridial cells to produce botulinum toxoid to be assessed. At this stage, therefore, the relative merits of the system, compared to other bacterial and eucaryotic expression systems, remain unknown.

REFERENCES

Abraham, L.J. & Rood J.I. (1985) *Plasmid 13,* 155-162.

Ahnert-Higler, G., Weller, U., Dauzenroth, M.E., Habermann, E. & Gratzl, M. (1989) *FEBS Microbiol. Lett. 242*, 245-248.

Allen, S.P. & Blaschek, H.P. (1988) *Appl. Environ. Microbiol. 54,* 2322-2324.

Blaustein, R. O., Germann, W. J., Finkelstein, A. & Dasgupta, B. R. (1987) *FEMS Microbiol. Letts. 226*, 115-120.

Binz, T., Kurazono, H., Wille, M., Frevert, J., Wernars, K. & Niemann (1990) *J. Biol. Chem. 265,* 9153-9158.

Binz, T., Kurazono, H., Popoff, M.R., Eklund, M.W., Sakaguchi, G., Kozaki, S., Krieglstein, K., Henschen, A., Gill, D.M. and Niemann, H. (1990) *Nucleic Acids Res. 18,* 5556.

Binz,T., O. Grebenstein, H. Kurazono, J. Thierer, U. Eisel, K. Wernars, M. Popoff, S. Mochida, B. Poulain, L. Tauc, S. Kozaki, and H. Niemann. (1991) Molecular biology of the L chains of clostridial neurotoxins. Abstracts of the 5th European Workshop on Bacterial Protein Toxins, Veldhoven, Holland, July 1991.

Bittner, M.A., DasGupta, B.R. & Holz, R.W. (1989) *J. Biol. Chem. 264*, 10354-10360.

Blasi, J., Chapman, E.R., Link, E., Binz, T., Yamasaki, S., de Camilli, P., Sudhof, T.C., Niemann, H. & Jahn J. (1993) *Nature 365*, 160-163.

Boquet, P. & Duflot, E. (1982) *Proc. Natl. Acad. Sci. USA 79*, 7614-7618.

Bruand, C., Ehrlich, S.D. & Janniere, L. (1991). *EMBO J. 10*, 2171-2177.

Brehm, J.K., Salmond, G.P.C. & Minton, N.P. (1987) *Nucl. Acids Res. 15*, 3177.

Brehm, J.K., Chambers, S.P., Bown, K.J., Atkinson, T. & Minton, N.P. (1992) *Appl. Microbiol. Biotechnol. 36,* 358-363.k

Campbell, K.D., Collins, M.D. & East, A.K. (1993) *J. Clin. Microbiol. 31*, 2255-2262.

Carter, P., Bedouelle, H. and Winter, G. (1985) *Nucleic Acids Res. 13*, 4431-4443.

Cato EP & Stackebrandt E (1989) in *Clostridia*, (Minton, N.P. & Clarke, D.J., eds), pp. 1-26, Plenum Publishing, New York.

Chambers S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. (1988) *Gene 68*, 139-149.

Chassy, B.M., Mercenier, A. & Flickinger, J. (1988) *Trends Biotechnol. 6*, 303-309.

Chen, E.Y. & Seeburg, P.H. (1985) *DNA 4*, 165-170.

Chou, P.Y. & Fassman, G.D. (1978) *Adv. Enzym. 47*, 45-

DasGupta, B.R. & Datta, A. (1987) *Fedn. Proc. Fedn. Am. Socs. Exp. Biol. 45*, 2289.

DasGupta, B.R. & Datta, A. (1988) *Biochimie (Paris) 70*, 811-817.

Dasgupta, B.R. (1990) *J. Physiol. (Paris) 84*, 220-228.

Dolly, J.O., Black, J., Williams, R.S. & Melling, J. (1984) *Nature 307*, 457-460.

East, A.K. & Dyke, K.G.H. (1989) *J. Gen. Microbiol. 135*, 1001-1015.

East, A.K., Richardson, P.T., Allaway, D., Collins, M.D., Roberts, T.A. & Thompson, D.E. (1992) *FEMS Microbiol. Letts. 96*, 225-230.

Eckert, K.A. & Kunkel, T.A. (1991) *PCR Methods and Applications 1*, 17-24.

Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J., Weller, U., Hudel, M., Habermann, E. & Niemann, H. (1986) *EMBO J. 5*, 2495-2502.

Evans, D.M., Williams, R.S., Shone, C.C., Hambleton, P., Melling, J., & Dolly, J.O. (1987) *Eur. J. Biochem. 154*, 409-416.

Fairweather, N.F., Lyness, V.A., Pickard, D.J., Allen, G. and Thomson, R.O. (1986) *J. Bacteriol. 165*, 21-27.

Fairweather, N.F., Lyness, V.A. and Maskell, D.J. (1987) I*nfect. Immunity 55*, 2541-2545.

Fairweather, N. F. & Lyness, V.(1986) *Nucleic Acid Res. 14*, 7809-7812.

Fujii, N., Kimura, K., Murakami, T., Indoh, I., Yashiki, T., Tsuzuki, K., Yokosawa, N. & Oguma, K. (1990) *Microbiol. Immunol. 34*, 1041-1047.

Garnier, J., Osguthorpe, D.J. & Robson, B. (1978) *J. Mol. Biol. 120*, 97-117

Garnier, T. & Cole, S.T. (1988) *Plasmid 19*, 134-150.

Gilbert H.J., Blazek, R, Bullman, H.M.S. & Minton N.P. (1986) *J. Gen. Microbiol. 132*, 151-160.

Gill, D.M. (1978) in *Bacterial toxins and cell membranes* (Jeljaszewicz, J. & Wadstrom, T., eds), pp. 291-332, Academic Press, Inc., New York.

Grosjean, H. & Fiers, W. (1982) *Gene 18*, 199-209.

Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acxad. Sci. USA 72*, 3961-3965.

Habermann, E. & Dreyer, F. (1986) *Curr. Top. Microbiol. Immunol. 129*, 93-179.

Hall, J.D., McCroskey, L., Pincomb, B.J. & Hatheway, C.L. (1985) *J. Clin. Microbiol. 21*, 654-655

Hatheway, C.L., McCroskey, L.M., Lombard, G.L. & Dowell, Jr., V.R. (1981) *J. Clin. Microbiol. 14*, 607-611.

Hauser, D., Eklund, M.W., Kurazono, H., Binz, T., Niemann, H., Gill, D.M. Boquet, P & Popoff, M.R. (1990) *Nucleic Acids Res. 18*, 4924.

Higgins, D.G. & Sharp, P.M. (1988) *Gene 73*, 237-244.

Hoch, D. H., Romero-Mira, M., Ehrlich, B. E., Finkelstein, A., Dasgupta, B. R. & Simpson, L. L. (1985) *Proc. Natl. Acad. Sci. USA 82*, 1692-1696.

Holmes, D.S., and Quigley, M. (1981) *Anal. Biochem. 114*, 193-197.

Kozaki, S., Kamata, Y., Takafumi, N., Ogasawara, J. & Sakaguchi, G. (1986) *Infect. Immun. 52*, 786-791.

Kozaki, S., Ogaswara, J., Shimote, Y., Kamata, Y. & Sakaguchi, G. (1987) *Infect. Immun. 55*, 3051-3056.

Kozaki, S., Miki, A., Kamata, Y., Ogasawara, J. & Sakaguchi, G. (1989) *Infect. Immun. 57*, 2634-2639.

Kyte, J. & Doolittle R.F. (1982) *J. Mol. Biol. 157*, 105-132.

Le Grice, S.F.J. Beuck, V. & Mous, J. (1987) *Gene 55*, 95-103.

Lucansky, J.B., Muriana, P.M. & Klaenhammer, T.R. (1988) *Mol. Microbiol. 2*, 637-646.

Maisey, E. A., Wadsworth, J. D. F., Poulain, B., Shone, C. C., Melling, J., Gibbs, P., Tauc, L. & Dolly, J. O. (1988) *Eur. J. Biochem. 177*, 683-691.

Marmur, J. (1961) *J. Mol. Biol. 3*, 208-218.

McCroskey, L.M., Hatheway, C.L. Fenicia, L., Pasolini, B. & Aureli, B. (1985) *J. Clin. Microbiol. 23*, 210-202.

McDonnell, M.W., Simon, M.N. & Studier, F.W. (1977) *J. Mol. Biol. 110*, 119-146.

Mead, D.A., Pey, N.K., Herrbstadt, C., Marcil, R.A. & Smith, L.M. (1991) *Biotechnol. 9*, 657-663.

Minton, N.P. (1984) *Gene 31*, 269-273.

Minton, N.P. & Morris, J.G. (1981) *J. Gen. Microbiol. 127*, 325-331.

Minton, N.P., Atkinson, T. & Sherwood, R.F. (1983a) *J. Bacteriol. 156*, 1222-1227.

Minton, N.P., Marsch, J.L. & Atkinson, T. (1983b) *J. Appl. Bacteriol. 55*, 445-452.

Minton, N. P., Atkinson, T., Bruton, C. J. & Sherwood, R. F. (1984) *Gene 31*, 31-38.

Minton, N.P., Bullman, H.M.S., Scawen, M.D., Atkinson, T. & Gilbert, H.J.. (1986) *Gene 46*, 25-35.

Minton, N.P. & Oultram, J.D. (1988) *Microbiol. Sci. 5*, 310-315.

Minton, N.P., Brehm, J.K., Oultram, J.D., Swinfield, T.J. & Thompson, D.E. (1988) in *Anaerobes Today* (Hardie, J.M. & Borellio, S.P., eds.), pp. 125-134. John Wiley & Sons, Chichester, UK.

Minton, N.P., Brehm, J.K., Oultram, J.D., Swinfield, T-J., Schimming, S., Whelan, S.M., Thompson, D.E., Young, M. & Staudenbauer, W.L. (1990a). *In* "Proceedings of the 6th International Symposium on Genetics of Industrial Microorganisms" Vol. II (H. Heslot, J. Davies, J. Florent, L. Bobichon, G. Durand and L. Penasse, Eds.), pp. 759-770. Sociètè Francaise de Microbiologie, Strasbourg.

Minton, N.P., Brehm, J.K., Oultram, et al. (1990b) in *Clinical and Molecular Aspects of Anaerobes* Borriello, S.P., ed.) pp. 187-201, Wrightson Biomedical Publishing Ltd, Petersfield.

Mochida, S., Poulain, B., Weller, U., Habermann, E. & Tauc, L. (1989) *FEBS Microbiol. Lett. 253*, 47-51.

Morris, N.P., Consiglio, E., Kohn, L.D., Habig, W.H., Hardegree, M.C. & Helting, T.B. (1981) *J. Biol. Chem. 255*, 6071-6076.

Notermans, S., Kozaki, Y. Kamata U, & Sakaguchi, G. (1984) *J. Med. Sci. Biol. 37*, 138-140.

O'Brien, R.W., and Morris, J.G. (1971) *J. Gen. Microbiol.* 68, 307-318.

Oultram, J.D., Davies, A., and Young, M. (1987) *FEMS Microbiol. Lett.* 42, 113-119.

Oultram, J.D., Loughlin, M., Swinfield, T.J., Brehm, J.K., Thompson, D.E. & Minton, N.P. (1988a) *FEMS Microbiol. Letts. 56*, 83-88.

Oultram, J.D., Peck, H., Brehm, J.K., Thompson, D., Swinfield, T.J. & Minton, N.P. (1988b) *Mol. Gen. Genet. 214*, 177-179.

Oultram, J.D., Loughlin, M., Walmsley, R., Gunnery, S.M. & Minton, N.P. (1993) The nucleotide sequence of genes involved in the leucine biosynthetic pathway in *Clostridium pasteurianum. Sequence* [in press].

Poulain , B., Mochida, S., Wadsworth, J.D.F., Weller, U., Hogy, B., Habermann, E., Dolly, J.O. & Tauc, L. (1990) *J. Physiol. Paris. 84*, 247-261.

Poulain , B., Mochida, S., Weller, U., Hogy, B., Habermann, E., Wadsworth, J.D.F., Shone, C.C., Dolly, J.O. & Tauc, L. (1991) *J. Biol. Chem. 266*, 9580-9585.

Reed, K.C., & Mann, D.A. (1985) *Nucl. Acids Res. 13*, 7207-7221.

Roberts, I., Holmes, W.M. & Hylemon, P.B. (1986) *Appl. Environ. Microbiol. 52*, 197-199.

Rood, J.I. & Cole, S.T. (1991) *Microbiol. Rev. 55*, 621-648.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning. A laboratory manual, second edition. Cold Spring harbour, New York.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B.A. (1980) *J. Mol. Biol. 143*, 161-178.

Sathyamoorthy, V. & Dasgupta, B.R. (1985) *J. Biol. Chem. 260*, 10461-10466.

Schiavo, G., Rossetto, O., Santucci, A., DasGupta, B.R., & Montecucco, C. (1992) *Nature 359*, 832-835.

Schiavo, G., Shone, C.C., Rossetto, O., Alexander, F.C.G. & Montecucco, C. (1993) *J. Biol. Chem. 268*, 11516-11519.

Schmidt, J.J., Sathyamoorthy, V. & Dasgupta, B.R. (1985) *Arch. Biochem. Biophys. 238*, 544-548.

Shone, C.C. & Hambleton, P. (1989) in *Clostridia* (Minton, N.P. & Clarke, D.J., eds) pp. 265-291, Plenum Publishing Inc., New York.

Shone, C. C., Hambleton, P. & Melling, J. (1985) *Eur. J. Biochem. 151*, 75-82.

Shone, C. C., Hambleton P. & Melling, J. (1987) *Eur. J. Biochem. 167*, 175-180.

Singh, B.R., Wasacz, F.M., Strand, S., Jakobsen, R.J. & Dasgupta, B.R. (1990) *J. Prot. Chem. 9*, 705-713.

Simpson, L.L. (1981) *Pharmacol. Rev. 33*, 155-188.

Simpson, L.L. (1986) *Ann. Rev. Pharmacol. Toxicol. 26*, 427-453.

Simpson, L.L. (1989) in *Botulinum Neurotoxin and Tetanus Toxin*, (Simpson, L.L. ed.) pp. 153-178, Academic Press Inc., San Diego.

Sloan, J., Warner, T.A., Scott, P.T., Bannam, T.L., Berryman, D.I. & Rood, J.I. (1992) *Plasmid 27*, 207-219.

Smith, D.B. & Johnson, K.S. (1988) *Gene 67*, 31-40.

Soutschek-Bauer, E., Hartl, L., and Staudenbauer, W.L. (1985) *Biotechnol. Lett.* 7, 705-710.

Stechner, B., Weller, U., Habermann, E., Gratzl, M. & Ahnert-Higler, G. (1989) *FEBS Microbiol. Lett. 255*, 391-394.

Sugiyama, H. (1980) *Microbiol. Rev. 44*, 419-448.

Swinfield, T.J., Oultram, J.D., Thompson, D.E., Brehm, J.K. & Minton, N.P. (1990) *Gene 87*, 79-90.

Swinfield, T-J., Janniere, L., Ehrlich, S.D. & MInton, N.P. (1991) *Plasmid 26*, 209-221.

Tabor, S. & Richardson, C.C. (1987) *Proc. Natl. Acad. Sci. USA 84*, 4767-4771.

Thompson, D.E., Brehm, J.K., Oultram, J.D., Swinfield, T-J., Shone, C.C., Atkinson, T., Melling, J. & Minton, N.P. (1990) *Eur. J. Biochem. 189*, 73-81.

Thompson, D.E., Hutson, R.A., East, A.K., Allaway, D. , Collins, M.D. & Richardson, P.T. (1993) *FEMS Microbiol. Letts 108*, 175-182.

Trach, K., Chapman, J.W., Piggot, P., LeCoq, D. & Hoch, J.A. (1988) *J. Bacteriol. 170*, 4194-4208.

Trieu-Cuot, P., Carlier, C., Martin, P. & Courvalin, P. (1987) *FEMS Microbiol. Lett. 48*, 289-294.

von Eichel-Streiber, C., Suckau, D., Wachter, M. & Hadding, U. (1989) *J. Gen. Microbiol. 135*, 55-64.

Weickert, M.J. Chambliss G.H. and Sugiyama, H. (1986) *Appl. Environ. Microbiol. 51*, 52-56.

Welloner, H.H. (1982) *Rev. Physiol. Biochem. Pharmacol. 93*, 1-68.

Wernars, K. & Notermans, S. (1990) in *Bacterial Protein Toxins* (Rappuoli, R., Alouf, J.E., Falmagne, P., Fehrenbach, F.J., Freer, J., Gross, R., Jeljaszewicz, J., Montecucco, C., Tomasi, M., Wadstrom, T & Witholt, B., eds), pp. 113-114, Gustav Fischer Verlag, Stuuttgart.

Williams, D.R., Young, D.I., Oultram, J.D., Minton, N.P., and Young, M. (1990a) in *Clinical and molecular aspects of anaerobes* (Borriello, S.P., ed.), pp. 239-246, Wrightson Biomedical Publishing Ltd, Petersfield.

Williams, D.R., Young, D.I., and Young, M. (1990b) *J. Gen. Microbiol.* 136, 819-826.

Wright et al 1992

Yamada, M. & Saier, M. (1987) *J. Biol. Chem. 262*, 5455-5463 .

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene 33*, 103-119.

Young, M , Staudenbauer, W.L. & Minton, N.P. (1989a) in *Clostridia* (Minton, N.P. & Clarke, D.J., eds) pp. 63-103, Plenum Publishing Inc., New York.

Young, M., Minton, N. P. & Staudenbauer, W. L. (1989b) FEMS *Microbiol. Rev. 63*, 1-25.

PUBLICATIONS:

GRANT NO: DAMD17-90-Z-0033

TITLE: Physical Characterisation of *Clostridium botulinum* Neurotoxin Genes.

PRINCIPAL INVESTIGATOR: Dr NIGEL P. MINTON

PI ADDRESS: Head of Molecular Genetics, Division of Biotechnology, PHLS Centre for Applied Microbiology & Research, Porton Down, Salisbury, Wiltshire, SP4 OJG, UNITED KINGDOM.

Journals:

Whelan SM, Elmore MJ, Bodsworth NJ, Atkinson T and Minton NP. (1992) The complete amino acid sequence of the *Clostridium botulinum* type-E neurotoxin, derived by nucleotide-sequence analysis of the encoding gene. Eur. J. Biochem **204:** 657-667.

Whelan SM, Elmore MJ, Bodsworth NJ, Brehm JK, Atkinson T and Minton NP. (1992) Molecular cloning of the *Clostridium botulinum* structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence. Appl. Environ. Microbiol. **58:** 2345-2354.

Books:

Whelan SM, Elmore MJ, Bodsworth NJ, Atkinson T and Minton NP. Amino acid sequence of the *Clostridium botulinum* type E neurotoxin. In Bacterial Protein Toxins, Eds. Witholt, B, Alouf, JE, Boulnois, GJ, Cossart, P, Dijkstra, BW, Falmagne, P, Fehrenbach, FJ, Freer, J, Niemann, H, Rappuoli, R and Wadstrom, T. (1992) Gustave Fischer Verlag, Stuttgart, Germany, pp. 95-96.

Elmore MJ, Whelan SM, Bodsworth NJ, Brehm JK, Atkinson T and Minton NP. Amino acid sequence of the *Clostridium botulinum* type B neurotoxin. In Bacterial Protein Toxins, Eds. Witholt, B, Alouf, JE, Boulnois, GJ, Cossart, P, Dijkstra, BW, Falmagne, P, Fehrenbach, FJ, Freer, J, Niemann, H, Rappuoli, R and Wadstrom, T. (1992) Gustave Fischer Verlag, Stuttgart, Germany, pp. 97-98.

Meeting Abstracts:

Elmore MJ, Whelan SM, Bodsworth NJ, Brehm JK, Atkinson T and Minton NP. Amino acid sequence of the *Clostridium botulinum* type B neurotoxin. Abstracts of 5TH European Workshop on Bacterial Protein Toxins, Veldhoven, June 1991.

Whelan SM, Elmore MJ, Bodsworth NJ, Atkinson T and Minton NP. Amino acid sequence of the *Clostridium botulinum* type E neurotoxin. Abstracts of 5TH European Workshop on Bacterial Protein Toxins, Veldhoven, June 1991.

In Preparation:

Elmore MJ, Bodsworth NJ, Whelan SM, and Minton NP. (1994) Molecular cloning of the structural gene encoding the type F neurotoxin from the proteolytic *C. botulinum* strain Langeland, and determination of its entire nucleotide sequence.

Bodsworth, NJ, Whelan SM and Minton NP (1994). Cloning and nucleotide sequence determination of the *C. botulinum* type G botulinum neurotoxin gene.

Mauchline ML and Minton NP (1994) Development of integrative procedures for *Clostridium acetobutylicum* NCIMB 8052.

Brehm JK, Mauchline ML and Minton NP (1994) Development of an expression vector for *Clostridium acetobutylicum* based on the promoter of the *Clostridium pasteurianum* ferredoxin gene.

PERSONNEL RECEIVING PAY

GRANT NO: DAMD17-90-Z-0033

TITLE: Physical Characterisation of *Clostridium botulinum* Neurotoxin Genes.

PRINCIPAL INVESTIGATOR: Dr NIGEL P. MINTON

PI ADDRESS: Head of Molecular Genetics, Division of Biotechnology, PHLS Centre for Applied Microbiology & Research, Porton Down, Salisbury, Wiltshire, SP4 OJG, UNITED KINGDOM.

Core Staff:

Ms N J Bodsworth
Dr M L Mauchline (nee Rodgers)

Supervisory:

Dr J D Oultram
Dr N P Minton (PI)

Undergraduate Assistants:
(Temporary)

Mr I Burr
Ms M Edwards